

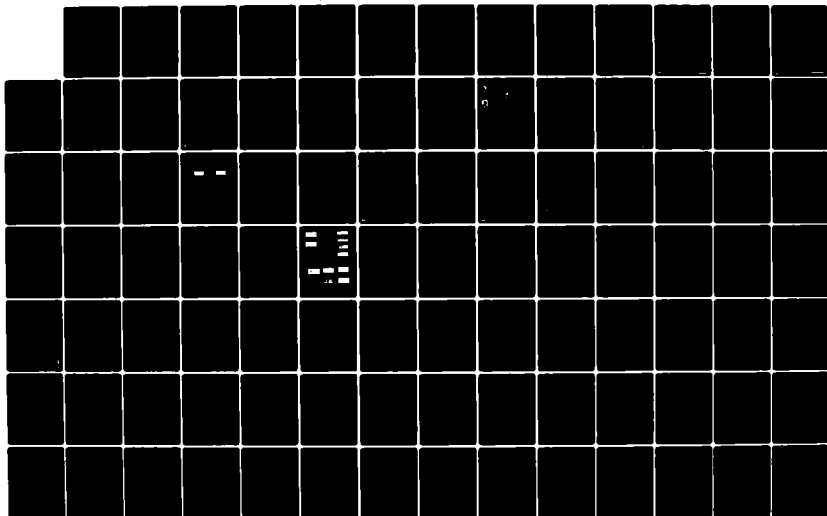
AD-A131-923

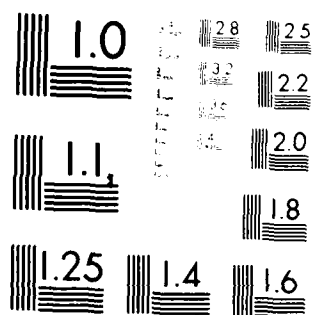
ACQUISITION IMAGE AND DATA COMPRESSION(U) CITY COLL NEW 1/2
YORK D L SCHILLING ET AL. 30 APR 83 AFOSR-TR-83-0505
AFOSR-81-0169

UNCLASSIFIED

F/G 17/2

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

13

AD A 131923

AFOSR-TR- 83 - 0505

ACQUISITION, IMAGE AND DATA
COMPRESSION
FINAL TECHNICAL REPORT
5444-2
1/81 - 2/28/83

Grant AFOSR-81-0169

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D. C. 20332

Donald L. Schilling
Professor
Co-Principal Investigator

George Eichmann
Professor
Co-Principal Investigator

COMMUNICATIONS SYSTEMS LABORATORY
DEPARTMENT OF ELECTRICAL ENGINEERING

DTIC

AUG 29 1983

Approved for public release;
distribution unlimited.



DTIC FILE COPY

THE CITY COLLEGE OF
THE CITY UNIVERSITY OF NEW YORK

83 06 21 08Y

ACQUISITION, IMAGE AND DATA COMPRESSION

Final Technical Report

5444-2

3/1/81 - 2/28/83

Grant AFOSR-81-0169

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D. C. 20332

Donald L. Schilling

Professor

Co-Principal Investigator

George Eichmann

Professor

Co-Principal Investigator

DEPARTMENT OF ELECTRICAL ENGINEERING

COMMUNICATIONS SYSTEMS LABORATORY

Accession For	
1	<input checked="" type="checkbox"/>
2	<input type="checkbox"/>
3	<input type="checkbox"/>
4	<input type="checkbox"/>
5	<input type="checkbox"/>
6	<input type="checkbox"/>
7	<input type="checkbox"/>
8	<input type="checkbox"/>
9	<input type="checkbox"/>
10	<input type="checkbox"/>
11	<input type="checkbox"/>
12	<input type="checkbox"/>
13	<input type="checkbox"/>
14	<input type="checkbox"/>
15	<input type="checkbox"/>
16	<input type="checkbox"/>
17	<input type="checkbox"/>
18	<input type="checkbox"/>
19	<input type="checkbox"/>
20	<input type="checkbox"/>
21	<input type="checkbox"/>
22	<input type="checkbox"/>
23	<input type="checkbox"/>
24	<input type="checkbox"/>
25	<input type="checkbox"/>
26	<input type="checkbox"/>
27	<input type="checkbox"/>
28	<input type="checkbox"/>
29	<input type="checkbox"/>
30	<input type="checkbox"/>
31	<input type="checkbox"/>
32	<input type="checkbox"/>
33	<input type="checkbox"/>
34	<input type="checkbox"/>
35	<input type="checkbox"/>
36	<input type="checkbox"/>
37	<input type="checkbox"/>
38	<input type="checkbox"/>
39	<input type="checkbox"/>
40	<input type="checkbox"/>
41	<input type="checkbox"/>
42	<input type="checkbox"/>
43	<input type="checkbox"/>
44	<input type="checkbox"/>
45	<input type="checkbox"/>
46	<input type="checkbox"/>
47	<input type="checkbox"/>
48	<input type="checkbox"/>
49	<input type="checkbox"/>
50	<input type="checkbox"/>
51	<input type="checkbox"/>
52	<input type="checkbox"/>
53	<input type="checkbox"/>
54	<input type="checkbox"/>
55	<input type="checkbox"/>
56	<input type="checkbox"/>
57	<input type="checkbox"/>
58	<input type="checkbox"/>
59	<input type="checkbox"/>
60	<input type="checkbox"/>
61	<input type="checkbox"/>
62	<input type="checkbox"/>
63	<input type="checkbox"/>
64	<input type="checkbox"/>
65	<input type="checkbox"/>
66	<input type="checkbox"/>
67	<input type="checkbox"/>
68	<input type="checkbox"/>
69	<input type="checkbox"/>
70	<input type="checkbox"/>
71	<input type="checkbox"/>
72	<input type="checkbox"/>
73	<input type="checkbox"/>
74	<input type="checkbox"/>
75	<input type="checkbox"/>
76	<input type="checkbox"/>
77	<input type="checkbox"/>
78	<input type="checkbox"/>
79	<input type="checkbox"/>
80	<input type="checkbox"/>
81	<input type="checkbox"/>
82	<input type="checkbox"/>
83	<input type="checkbox"/>
84	<input type="checkbox"/>
85	<input type="checkbox"/>
86	<input type="checkbox"/>
87	<input type="checkbox"/>
88	<input type="checkbox"/>
89	<input type="checkbox"/>
90	<input type="checkbox"/>
91	<input type="checkbox"/>
92	<input type="checkbox"/>
93	<input type="checkbox"/>
94	<input type="checkbox"/>
95	<input type="checkbox"/>
96	<input type="checkbox"/>
97	<input type="checkbox"/>
98	<input type="checkbox"/>
99	<input type="checkbox"/>
100	<input type="checkbox"/>



AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Bolling Air Force Base
Washington, D. C. 20332
Distribution
MAILED J. 8
Chief, Technical Information Division

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR. 83-0505	2. GOVT ACCESSION NO. AD A131 523	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ACQUISITION, IMAGE AND DATA COMPRESSION		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report 3/1/81 - 2/28/83
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Donald L. Schilling George Eichmann		8. CONTRACT OR GRANT NUMBER(s) AFOSR-81-0169
9. PERFORMING ORGANIZATION NAME AND ADDRESS City College of New York New York, N. Y. 10031		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2305/B1
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling, A.F.B., Washington, D. C. 20332		12. REPORT DATE April 30, 1983
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES 102
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
15a. DECLASSIFICATION DOWNGRADING SCHEDULE		
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Presentations made at MILCOM'82; Optical Society '83; SPIE'83; Optical Computing Conference'83.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Spread Spectrum, Optical Transforms, Acquisition, Simulation, Tracking, PN Direct Sequence Frequency Hopping		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report discusses the following subjects: A new technique to achieve a fast PN Acquisition scheme. A comparison of schemes for coarse acquisition of frequency-hopped spread-spectrum signals; Investigation of the tracking of frequency hopped spread spectrum signals in adverse environments.		

UNCLASSIFIED

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

· Coherent optical production of the Hough transform.

Estimation of the closely spaced frequencies buried in white noise using linear programming.

Restoration of discrete Fourier spectra using linear programming.

Two-dimensional optical filtering of 1-D signals.

Random TDMA Access Protocol with application to multi beam satellites.

A reservations scheme of multiple access for local networks.

A random TDMA access protocol for satellite channel has been analyzed. The analysis is based on obtaining the idle and busy period distribution. The average packet queueing delay in the user's buffer is obtained. A modified random TDMA is suggested to improve the performance. Implementation of the scheme in a multi-spot beam satellite is presented and the delay on board the satellite is obtained.

A reservation scheme for multiple access suited for local networks of broadcast channels is considered and analyzed. The analysis is based on two imbedded Markov chain models, one for the individual user and one for the entire set of users. The average queue size as well as the average delay time are obtained.

A fast algorithm for scene matching is studied and compared with other algorithms. The measure of similarity is the sum of the absolute values of the differences of image intensity between the template and the scene. Comparison with other measures of similarity is presented.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Table of Contents

	<u>Page</u>
1.0 Research Objectives	1
1.1 A Fast PN Acquisition Scheme	1
1.2 Tracking of Frequency Hopped Spread Spectrum Signals in Adverse Environments	1
1.3 Comparison of Schemes for Coarse Acquisition of FH Spread Spectrum Signals	1
1.4 Two-Dimensional Optical Filtering of 1-D Signals	1
1.5 Coherent Optical Production of the Hough Transform	2
1.6 Estimation of Two Closely Spaced Frequencies	2
1.7 Restoration of Discrete Fourier Spectra	2
1.8 Study of a Random Access Scheme for Multi-Beam Satellites	2
 2.0 Status of the Research Effort	 3
2.1 A Fast PN Acquisition Scheme	4
2.2 A Comparison of Schemes for Coarse Acquisition of Frequency-Hopped Spread-Spectrum Signals	9
2.3 Tracking of Frequency Hopped Spread Spectrum Signals in Adverse Environments	16
2.4 Coherent Optical Production of the Hough Transform	22
2.5 Estimation of Two Closely Spaced Frequencies Buried in White Noise Using Linear Programming	27
2.6 Restoration of Discrete Fourier Spectra Using Linear Programming	31

Table of Contents (Continued)

	<u>Page</u>
2.7 Two-Dimensional Optical Filtering of 1-D Signals	37
2.8 Random TDMA Access Protocol with Application to Multi Beam Satellites	42
2.9 A Reservations Scheme of Multiple Access for Local Networks	50
2.10 An Algorithm for Scene Matching	77

1.0 RESEARCH OBJECTIVES

1.1 A Fast PN Acquisition Scheme

In this project we studied acquisition techniques for DS spread spectrum signals. The technique utilized two thresholds rather than a single threshold. As a result, the time to acquire was reduced by an order of magnitude as compared to the techniques proposed by other investigators who use single threshold techniques. The procedure can be employed with most acquisition techniques merely by replacing the single threshold with the two threshold system.

1.2 Tracking of Frequency Hopped Spread Spectrum Signals in Adverse Environments

The objective of this research is to design tracking strategies in adverse environments such as fading and jamming. This led to optimization of tracking in the sense of maximizing the mean time to loss of lock.

1.3 Comparison of Schemes for Coarse Acquisition of FH Spread Spectrum Signals

The objective of this research was to compare three of the more popular acquisition techniques on the basis of probability of miss, time to acquire and complexity.

1.4 Two-Dimensional Optical Filtering of 1-D Signals

The object of this study was to implement a generalized space-spatial frequency filter and study its operation.

1.5 Coherent Optical Production of the Hough Transform

The objective of this research was to study the operation of the forward Radon transform (FRT) and compare it to the standard Fourier transform.

1.6 Estimation of two Closely Spaced Frequencies

The objective of this research was to use linear programming techniques to improve the resolution of a spectrum analyzer.

1.7 Restoration of Discrete Fourier Spectra

The objective of this study is to examine the use of linear programming for superresolution. A technique is presented which greatly enhances the resolution of a spectrum analyzer.

1.8 Study of Random Access Scheme for Multi-Beam Satellite

The objective of this study is to examine the study of a reservation access scheme for local area networks. We also examine the study of an algorithm for scene matching, using the sum of the absolute values of the differences of image intensity between the template and the scene as the measure of similarity.

2. Status of the Research Effort

A FAST PN ACQUISITION SCHEME

Sorin Davidovici

Laurence B. Milstein

Donald L. Schilling

TRW
San Diego, CaliforniaDepartment of Elec. Eng. & Computer Science
University of California, San Diego
La Jolla, CaliforniaDept. of Elec. Eng.
City College of New York
New York, New York

ABSTRACT

This paper describes a new technique to achieve rapid acquisition in a direct sequence spread spectrum communication system. It is based in part on an upper bound for the partial correlation function of a PN sequence, and hence it is a fairly general procedure, applicable to a wide variety of systems.

INTRODUCTION

The problem of acquisition of a spread spectrum signal consists of correctly estimating the phase of a received PN sequence of arbitrary length $L = 2^N - 1$, where N refers to the number of delay stages in the generating shift register or, equivalently, to the degree of the generating polynomial which describes the feedback connections. The acquisition process consists of repeatedly comparing the received PN sequence $g(t - jT_c)$, with a local generated PN sequence, $g(t - iT_c)$, until the phases iT_c and jT_c are found to be equal. At that point the locally generated PN sequence is said to be synchronized to the received PN sequence and the acquisition process is ended. Throughout this paper, the carrier phase and chip timing are assumed to be in perfect synchronism. The in-phase/not in-phase decision is based on a partial correlation process. As is well known, the partial correlation function of a PN sequence is not nearly as well behaved as the autocorrelation function, it being a function of the particular starting phases i and j , the specific feedback connections, and the length of the correlation γT_c , where γ equals the number of chips (an integer) and T_c equals the chip duration. To circumvent all of these difficulties, various bounds have been found ([1]-[3]) to the partial autocorrelation function, and an approximation to one of these bounds is used in the analysis presented in the next section.

ANALYSIS

Reference [2] describes a recursive method for generating a non-linear sequence which serves as an upper bound to the partial autocorrelation function of a PN sequence (via the shift-and-add property). This bound, unfortunately, is defined via a recursive algorithm as opposed to an analytical form. It can be shown that a good approximation to this bound is given by $\gamma(1 - \gamma/L)$, and in what follows, we will use this approximation in order to obtain a simple analytical result. Using the approximate bound, it follows that for $i \neq j$,

$$\int_0^{\gamma T_c} g(t - jT_c) g(t - iT_c) dt \approx T_c \gamma \left(1 - \frac{\gamma}{L}\right) \quad (1)$$

The output of the correlator in Fig. 1 is

$$V_o(\gamma T_c) = \int_0^{\gamma T_c} [\sqrt{2P_s} g(t - jT_c) + n_w(t)] g(t - iT_c) dt \quad (2)$$

which consists of a signal term $S_o(\gamma T_c)$ and a noise term $n_o(\gamma T_c)$. The noise term can easily be shown to be Gaussian with zero mean and variance $\sigma_o^2(\gamma T_c) = \frac{\gamma}{2} \gamma T_c$, where $\frac{\gamma}{2}$ is the two-sided power spectral density of the input noise. The signal term, $S_o(\gamma T_c)$ is given by

$$\begin{aligned} S_o(\gamma T_c) &= \int_0^{\gamma T_c} \sqrt{2P_s} g(t - jT_c) g(t - iT_c) dt \\ &= \sqrt{2P_s} \gamma T_c \quad \text{if } i = j \\ &\approx \sqrt{2P_s} \gamma T_c \left(1 - \frac{\gamma}{L}\right) \quad \text{if } i \neq j \end{aligned} \quad (3)$$

The synchronism/no synchronism decision is seen only to involve the choice of which signal term, $S_o(\gamma T_c)$, is present at the correlator's output. Due to the presence of noise, this decision can only be made with a specified accuracy. If as shown in Fig. 2, we wait for a time $T = \gamma T_c$ before making a decision, then we can define two voltage regions such that if the correlator's output falls in region I, we decide synchronism is attained, and if we fall in region II, we decide the two PN sequences are not synchronized. The probability of an erroneous decision, given we are not looking at the correct phase position, is equal to

$$\begin{aligned} P_e &= P[V_o(\gamma T_c) > V_T] \\ &= P[n_o(\gamma T_c) > c_1] \end{aligned} \quad (4)$$

where c_1 is shown in Fig. 2 and V_T , the threshold, is given below [see(7)].

Note that both in Fig. 2 and from Eq. (4), for simplicity the decision boundary has been located halfway between the two signals. Therefore,

$$\epsilon_1 = \frac{1}{2} \left[\sqrt{2P_s} \gamma_1 T_c - \sqrt{2P_s} \gamma_1 T_c \left(1 - \frac{\gamma_1}{L} \right) \right] \quad (5)$$

$$= \frac{1}{2L} \sqrt{2P_s} \gamma_1^2 T_c$$

The probability of error of Eq. (4) then becomes

$$P_e = P[n_o(\gamma_1 T_c) > \frac{1}{2L} \sqrt{2P_s} \gamma_1^2 T_c] \quad (6)$$

$$= \frac{1}{2} \operatorname{erfc} \left[\frac{\gamma_1}{L} \sqrt{\frac{E_{s1}}{2\eta}} \right]$$

where $E_{s1} = \gamma_1 E_c$, E_c is the energy per chip and equals $P_s T_c$, and

$$\operatorname{erfc}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-y^2} dy$$

Finally, $V_T(\gamma_1 T_c)$ is given by

$$V_T(\gamma_1 T_c) = \sqrt{2P_s} \gamma_1 T_c - \epsilon_1$$

$$= \sqrt{2P_s} \gamma_1 T_c \left[1 - \frac{\gamma_1}{2L} \right] \quad (7)$$

In what follows, we will show how the acquisition process can be improved by shortening the acquisition time.

The Decision Process Prior to $\gamma_1 = \gamma_1$.

At times it may not be necessary to wait the full γ_1 chips to make a decision. The only requirement is that at any time a decision is made, the probability of error, P_e , be kept constant. If any attempt at making a decision after, say, γ_2 chips is made, then the threshold voltage must be set such that

$$P_e = P[n_o(\gamma_2 T_c) > \epsilon_2] = \frac{1}{2} \operatorname{erfc} \left[\frac{\epsilon_2}{\sqrt{2\sigma_o}(\gamma_2 T_c)} \right] \quad (8)$$

Note that Eq. (8) is identical to Eq. (4) with the exception of the pertinent parameters (i.e. the decision mechanism is identical). But if the probability of error, P_e , is to be kept constant, then

$$P_e|_{\text{at } t = \gamma_1 T_c} = P_e|_{\text{at } t = \gamma_2 T_c}$$

or

$$\frac{1}{2} \operatorname{erfc} \left[\frac{\epsilon_1}{\sqrt{2\sigma_o}(\gamma_1 T_c)} \right] = \frac{1}{2} \operatorname{erfc} \left[\frac{\epsilon_2}{\sqrt{2\sigma_o}(\gamma_2 T_c)} \right] \quad (9)$$

This implies that the two arguments of the erfc function must be equal. Hence

$$\frac{\epsilon_1}{\sqrt{2\sigma_o}(\gamma_1 T_c)} = \frac{\epsilon_2}{\sqrt{2\sigma_o}(\gamma_2 T_c)}$$

or,

$$\epsilon_2 = \epsilon_1 \frac{\sigma_o(\gamma_2 T_c)}{\sigma_o(\gamma_1 T_c)} = \frac{1}{2L} \sqrt{2P_s} \frac{\gamma_2}{\gamma_1} \gamma_1^2 T_c \quad (10)$$

Eq. (10) defines the parameter ϵ_2 such that at any time $T = \gamma_2 T_c$ a decision can be attempted. Fig. 2 shows qualitatively the decision regions at any given $\gamma_2 < \gamma_1$.

If at time $T = \gamma_2 T_c$ the correlator's output voltage equals V_o , the following decisions can be made with a given probability of error P_e :

- 1) Synchronization has been attained ($i = j$) if

$$V_o > V_s = \sqrt{2P_s} \gamma_2 T_c \left(1 - \frac{\gamma_2}{L} \right) + \epsilon_2$$

- 2) Synchronization will not be attained ($i \neq j$) if

$$V_o < V_n = \sqrt{2P_s} \gamma_2 T_c - \epsilon_2$$

- 3) No decision can be made; continue the correlation if

$$V_n < V_o < V_s$$

Thus, $V_n(\gamma_2)$ and $V_s(\gamma_2)$ become

$$V_n(\gamma_2) = \sqrt{2P_s} \gamma_2 T_c - \epsilon_2$$

$$= \sqrt{\frac{P_s}{2}} \gamma_1 T_c \left[2 \frac{\gamma_2}{\gamma_1} - \frac{\gamma_1}{L} \sqrt{\frac{\gamma_2}{\gamma_1}} \right] \quad (11)$$

and

$$V_s(\gamma_2) = \sqrt{2P_s} \gamma_2 \left(1 - \frac{\gamma_2}{L} \right) + \epsilon_2$$

$$= \sqrt{\frac{P_s}{2}} \gamma_1 T_c \left[2 \frac{\gamma_2}{\gamma_1} \left(1 - \frac{\gamma_2}{L} \right) + \frac{\gamma_1}{L} \sqrt{\frac{\gamma_2}{\gamma_1}} \right] \quad (12)$$

From Fig. 2 the probability of no decision can be seen to equal $P(\text{no decision})$

$$\triangleq P_{nd} = P[V_n < s_o + n_o < V_s] \quad (13)$$

With V_n and V_s given in equations (11) and (12), respectively, the probability of no decision at time $\gamma_2 T_c$ becomes

$$P_{nd_2} = \frac{1}{2} \left[\operatorname{erf} \left[\sqrt{\frac{E_s}{2n}} \left(\frac{\gamma_1}{L} - 2 \sqrt{\frac{\gamma_2}{\gamma_1} \frac{\gamma_2}{L}} \right) \right] + \operatorname{erf} \left[\sqrt{\frac{E_s}{2n}} \frac{\gamma_1}{L} \right] \right]$$

If Q_{d_2} is defined as the probability of making a decision at $T = \gamma_2 T_c$, then

$$Q_{d_2} = 1 - P_{nd_2}$$

The realization that a decision can be made prior to $t = \gamma_1 T$ without compromising the system's error performance can be shown to decrease the PN acquisition time. A detailed analysis of the actual improvement in the acquisition performance obtained by the application of this technique is rather lengthy and is given in [4].

NUMERICAL RESULTS

Figures (3a) to (3f) show curves of probability of decision versus the chip number (i.e. the value of γ_2) for various sets of system parameters. The ratio of energy-per-chip-to-noise spectral density E/n has a value of -10dB, -5dB or 0dB, and the probability of making an error is taken to be either 0.1 or 0.01. The period of the PN sequence is 1023 (i.e., $N = 10$) in all the figures.

On each figure there are two curves, one showing the probability of making a decision at or before any time T , and the other showing the conditional probability of making the decision at any specific time T (i.e. Q_{d_2}). It can be seen from the figures that there is a high probability that a decision can be made before time γ_1 with no loss of system performance.

REFERENCES

1. S. Davidovici and D. L. Schilling, "Minimum Acquisition Time of a PN Sequence," NTC '78, pp. 35.6.1-35.6.4.
2. S. Davidovici, G. Sevaston, and D. L. Schilling, "PN Sequence Acquisition," NTC '79, pp 54.5.1-54.5.4.
3. F. Hemmati and D. L. Schilling, "Worst Case Acquisition of PN Sequences." Submitted to IEEE Trans. Comm.
4. S. Davidovici, L. B. Milstein, and D. L. Schilling, "A New Rapid Acquisition Technique for DS Spread Spectrum Communications." To be submitted to IEEE Trans. Comm.

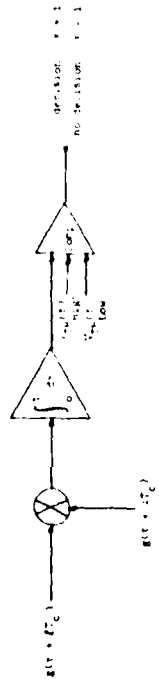
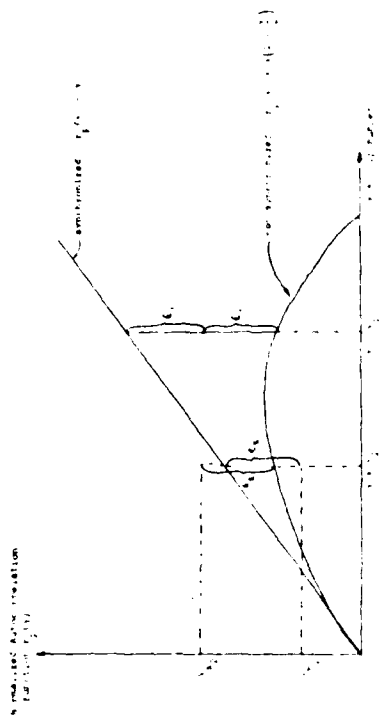
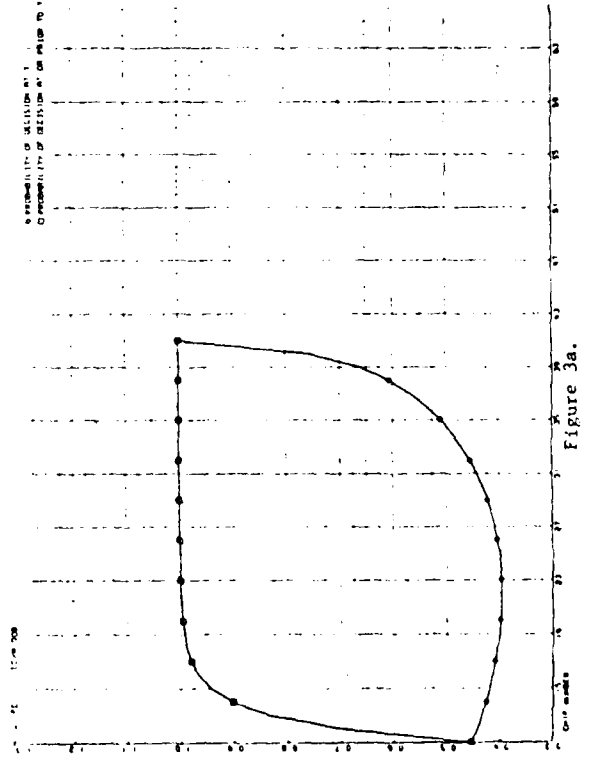
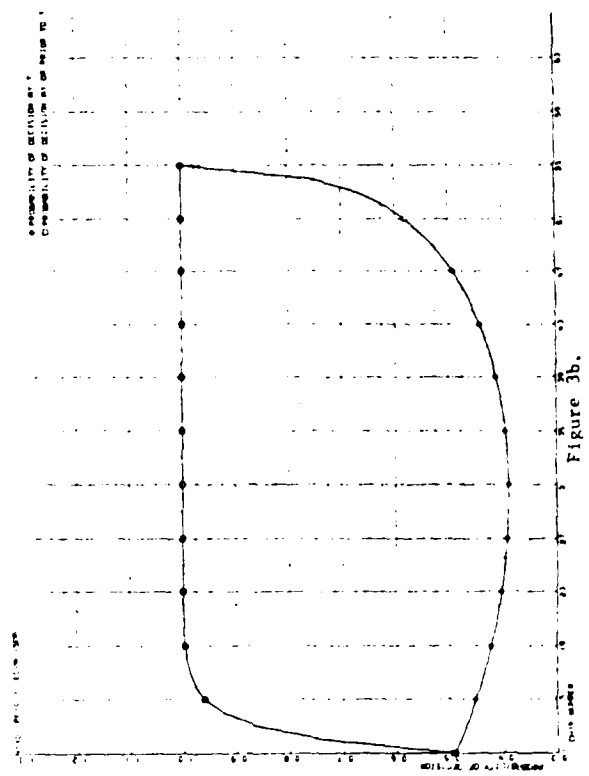


Figure 3.



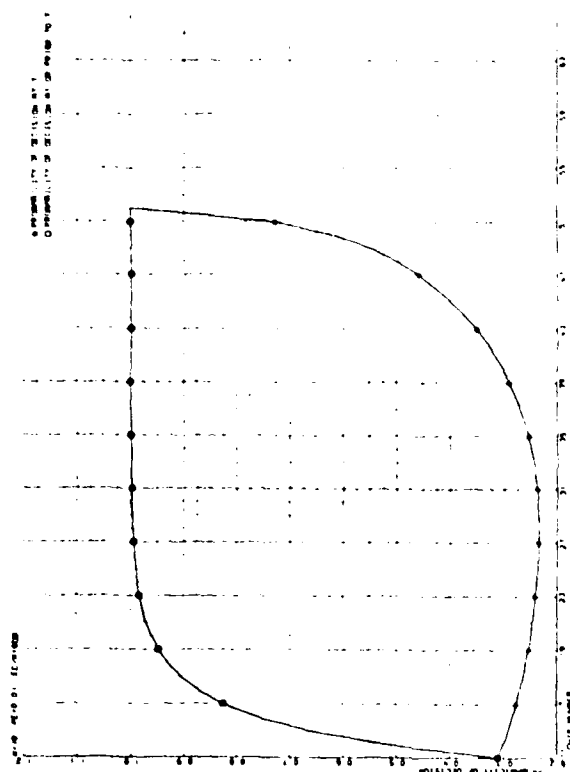


Figure 3d.

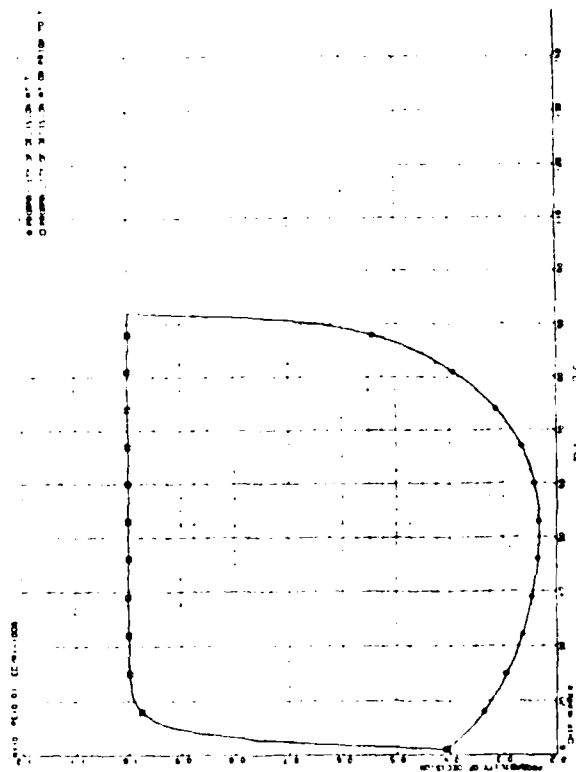


Figure 3e.

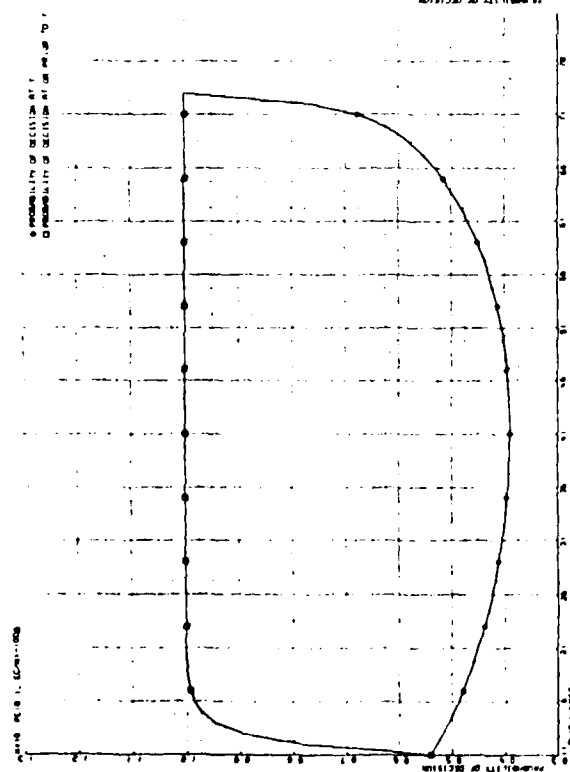


Figure 3c.

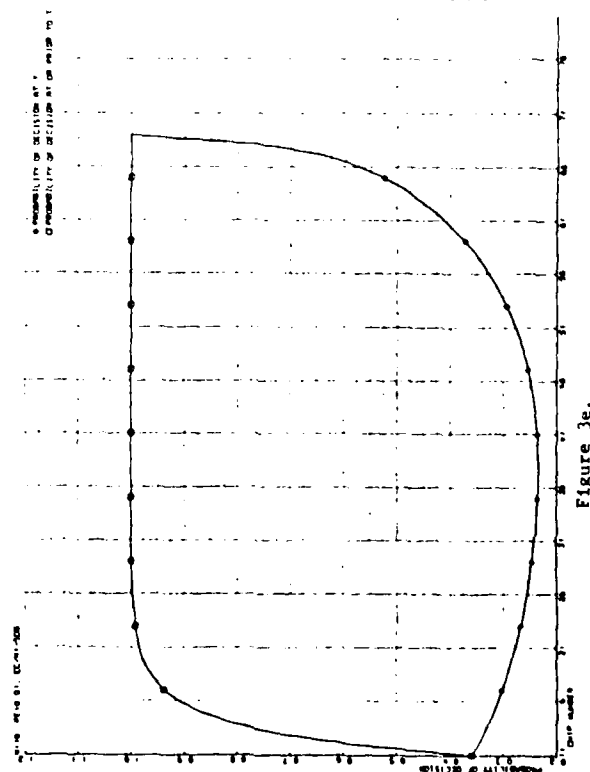


Figure 3b.

A Comparison of Schemes for Coarse Acquisition of Frequency-Hopped Spread-Spectrum Signals

CLIVE A. PUTMAN, STEPHEN S. RAPPAPORT, SENIOR MEMBER, IEEE, AND DONALD L. SCHILLING, FELLOW, IEEE

Abstract—Three schemes for coarse acquisition of spread-spectrum signals are compared: stepped serial search, matched filter, and two-level. Analytical models and formulas are developed which characterize performance in adverse environments. Comparisons are on the basis of miss probability, mean time to search uncertainty region, and relative complexity.

I. INTRODUCTION

THE synchronization process required for spread-spectrum systems consists of two parts—acquisition and tracking. This paper describes and compares several methods for acquisition. Three schemes are considered: stepped serial search, matched filter, and two-level acquisition. These schemes have been well described in the technical literature [1]–[4]. The intention here is to develop analytical models and mathematical formulas which characterize the performance of the schemes in adverse environments, and to compare the schemes in terms of miss probability, uncertainty region search time, and relative complexity.

The investigation is primarily motivated by consideration of ground radio frequency-hopping schemes. It is assumed that communicators operate in a push-to-talk mode and that a synchronization process is initiated with each transmission.

A brief description of the schemes under consideration is followed by a general derivation of performance measures. Models for Rician fading and user or jammer interference environments are characterized and used to predict performance. Methods used in obtaining numerical results are outlined and the resulting performance curves are discussed with reference to particular spread-spectrum applications.

II. GENERAL DESCRIPTION

In the stepped serial acquisition scheme shown diagrammatically in Fig. 1, the timing epoch of the local PN code is set and the locally generated signal correlated with the incoming signal. If, at the end of an examination interval, the threshold is not exceeded, the search control inhibits a clock pulse to the PN code generator so that the local code phase slips to the next cell, n cells per chip, and the process is then repeated.

Paper approved by the Editor for Communication Systems Disciplines of the IEEE Communications Society for publication after presentation at the International Conference on Communications, Denver, CO, June 1981. Manuscript received May 30, 1981; revised July 20, 1982. This work was supported in part by S Consulting Services (now SCS Telecom, Inc.), Sands Point, NY.

C. A. Putman and D. L. Schilling are with the Department of Electrical Engineering, City College of New York, New York, NY 10031.

S. S. Rappaport is with the Department of Electrical Engineering, State University of New York at Stony Brook, Stony Brook, NY 11794.

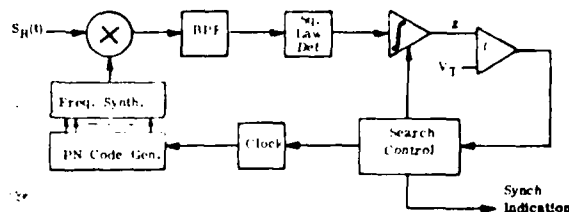


Fig. 1. Stepped serial search scheme.

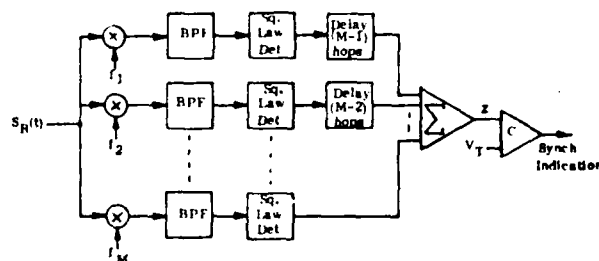


Fig. 2. Matched filter scheme.

The system thus performs an active correlation with a received code sequence.

The matched filter, a passive correlator, searches the incoming code in real time. The essential structure is shown in Fig. 2. A sequence of M consecutive frequencies is chosen by the receiver to establish a code start epoch and the matched filter performs a near optimal noncoherent detection as the sequence is received.

The two-level scheme combines passive and active correlation and, as such, combines the capability of searching the code in real time with integration over a large number of chips. As shown in Fig. 3, a matched filter detects a relatively short M hop synch prefix and applies code start signals to a bank of c active correlators. Each signal causes the next idle correlator to cycle through K hops at the end of which any correlator output exceeding the second threshold causes a synch indication; otherwise, the correlator is again made available to the common bank.

III. PERFORMANCE MEASURES

Detection Probabilities

For ground radio systems operating in a push-to-talk mode the prime concern is to establish the desired signal code phase during some brief initial acquisition period. We thus characterize the performance of an acquisition scheme in terms of the probability that, at the end of the examination interval, the desired signal code sequence is not detected. This is a func-

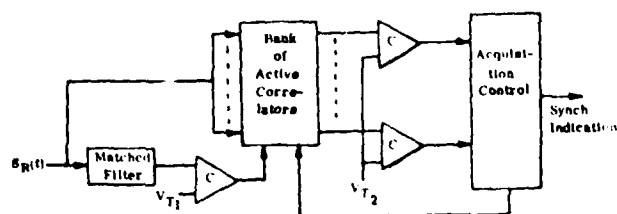


Fig. 3. Two-level scheme.

tion of the threshold level setting, and for the purposes of comparing schemes we set this parameter by specifying the probability that the threshold is exceeded when the desired signal code sequence is not present. The performance measure is then the miss probability for a given false alarm probability.

The input waveform in the i th hop interval can be characterized as a narrow-band waveform

$$S_R(t) = A_i \cos [w_i(t) + \theta_i] + g_i(t) \quad (1)$$

where w_i is determined by a known pseudorandom sequence, and $g_i(t)$ is a Gaussian process with variance σ^2 . This signal can be expressed as

$$S_R(t) = r_i(t) \cos [w_i(t) + \theta_i + \phi_i(t)] \quad (2)$$

where $r_i(t)$ is a Rician distributed process.

In the case of an active correlator the local waveform is hopped at the same rate as the input signal and the detector output is integrated over M hop periods. In the case of a passive correlator the outputs from M similar mixers and detectors for successive frequencies are stored and added. The system thus forms the sum of squares of independent Rician variates and the output Z can be shown [5] to be noncentral chi-squared distributed with $2M$ degrees of freedom. For convenience we define a normalized variable Z_0 , such that $Z_0 = Z/\sigma^2$. Its first-order probability density is then given by

$$p_{Z_0}(z_0) = \frac{1}{2} (z_0/S)^{(M-1)/2} e^{-(z_0+S)/2} I_{M-1}(\sqrt{S}z_0) \quad (3)$$

where the noncentral parameter

$$S = \sum_{i=1}^M A_i^2 / \sigma^2$$

and I_{M-1} is the modified Bessel function of the first kind, order $M-1$.

If the output exceeds a threshold level V_T a synch detection is declared. The detection probability is then

$$P_D = P\{Z > V_T\} = \int_L^\infty p_{Z_0}(z_0) dz_0 \quad (4)$$

where $L = V_T/\sigma^2$. This can be expressed in terms of the generalized Marcum Q -function [6] as

$$P_D = Q_M(\sqrt{S}, \sqrt{L})$$

where

$$Q_M(a, b) = \int_b^\infty x(x/a)^{M-1} e^{-(a^2+x^2)/2} I_{M-1}(ax) dx.$$

Its complement is the miss probability expressed as

$$P_M = 1 - P_D = Q_M^c(\sqrt{S}, \sqrt{L}). \quad (6)$$

The false alarm probability is the probability that the threshold is exceeded when the correct code sequence is not present. In the case of the benign environment (noise only—no fading or interferers) considered thus far, the false alarm probability is

$$P_{FA} = Q_M(0, \sqrt{L}). \quad (7)$$

Search Time

It is to be emphasized that the acquisition scheme is required to reliably detect the first occurrence of the correct code epoch during the search process. We do not consider systems which can tolerate one or more misses of the code epoch and continue to seek acquisition. Let the uncertainty region length N_c be the maximum expected delay between the locally generated and desired signal code phases measured in hop intervals. Thus, a second performance measure which must be considered is the time taken to search the uncertainty region, T_s .

IV. ENVIRONMENT MODELING

Fading Channel Model

We consider a Rician fading channel so that the received signal with which synchronization is desired is given during the i th hop interval by

$$S_R(t) = \sqrt{2a_i P_s} \cos(w_i t - \theta_i) + \sum_j c_j \sqrt{2P_s} \cos(w_j t - \theta_j) + n_w(t) \quad (9)$$

consisting of specular, scatter, and noise components, respectively. The scatter and noise components are Gaussian and independent and the scatter component is signal dependent. Equation (9) defines the fading model, and for $a_i = 0$, is identical to that given in [7].

For comparative analysis we equate the sum of powers in the specular and scatter components to a constant P_s , representing the average received signal power over many hops. In particular we take

$$\hat{a} + b = 1 \quad (10)$$

where

$$\hat{a} = \frac{1}{M} \sum_i a_i \quad \text{and} \quad b = \sum_j c_j^2.$$

(5) Here \hat{a} is summed over many hops and b is summed over many

multipaths. The quantities \hat{a} and b are indicative of the relative strength of the specular and scatter components, respectively. It is assumed that these parameters are fixed over the acquisition period, although they may have long-term slow variations.

Following a derivation similar to that in [7], the above signal can be expressed in the form of (2). The Rician distributed envelope at the input to the detector now has parameters

$$A_i \approx \sqrt{2\hat{a}P_s}$$

and

$$\sigma^2 = N_0 B_{IF} + bP_s \quad (11)$$

where N_0 is the one-sided noise power spectral density and B_{IF} is the detector bandpass filter bandwidth.

Multiple Access Model

We now consider the interference generated by other users who are hopping over the set of F frequency slots. The probability that no other user causes interference during a given hop when N other users are active is $(1 - 1/F)^N$. The probability that at least one other user causes interference is

$$p = 1 - (1 - 1/F)^N \quad (12)$$

Since we are integrating over M hops, the probability that j slots have at least one user present is

$$b(j, M, p) = \binom{M}{j} p^j (1 - p)^{M-j} \quad (13)$$

Jamming Model

In application, the receiver could provide for limiting of received signal strength to slightly above that of the desired signal. This mitigates the effects of CW jamming. An optimized jammer would be forced to jam a random selection of frequencies. When the number of frequencies available F is large, the above model is reasonable for random hop or comb jamming if N is taken as the number of interferers or interfering tones, respectively.

The Combined Model

To quantify performance for a fading environment with interferers present it is analytically convenient to assume that the power of interfering signals lies in the specular components, and that hopped interfering signals are (hop) synchronous with the desired signal. In addition, all interferers are assumed to be of equal strength as "heard" at the receiver. Interferers are assumed to be statistically independent RF sources with random phase. Our approximate performance analysis proceeds by noting that the sum of interfering and desired sinusoids produces a resultant sinusoidal signal at the receiver. Because of the assumptions above, the average power in the resultant is the sum of the powers in the components, and we proceed by using this effective sinusoid as the resultant specular component at the receiver.

The false alarm probability given j slots are jammed is then

$$P(\text{false alarm} | j) = Q_M(\sqrt{S_j}, \sqrt{L}) \quad (14)$$

with

$$S_j = j2JP_s/N_0 B_{IF} = j2JE_c/N_0$$

where E_c is the received energy per hop, J is the jammer power (or limited signal power) to desired signal power ratio, and we assume that $B_{IF} = 1/T_h$, T_h being the hop interval. The overall false alarm probability is then determined by multiplying by the probability of j slots being jammed and summing over j ,

$$P_{FA} = \sum_{j=0}^M b(j, M, p) Q_M(\sqrt{S_j}, \sqrt{L}) \quad (15)$$

Similarly, the miss probability is

$$P_M = \sum_{j=0}^M b(j, M, p) Q_M^c(\sqrt{S_d}, \sqrt{L'}) \quad (16)$$

with

$$S_d = \frac{[j(J + \hat{a}) + (M - j)\hat{a}] 2P}{N_0 B_{IF} + bP_s} \quad s = \frac{[jJ + M\hat{a}] 2E_c/N_0}{1 + bE_c/N_0}$$

and

$$L' = L/(1 + bE_c/N_0).$$

V. ACQUISITION SCHEME PERFORMANCE

Stepped Serial Search

There are many ways to control the search process [8]. In the scheme considered here it is assumed that three successive threshold exceedences, or hits, are required to initiate the tracking phase. Any one miss causes the cell to be rejected. Hence, the overall probability of false alarm is

$$P_{FA} = P_{FH}^3 \quad (17)$$

where P_{FH} , the probability of a false hit, is given by (15), and the overall miss probability is

$$P_M = 1 - P_H^3 \quad (18)$$

where P_H is the probability of a hit.

The magnitude of the detected energy is dependent on the relative delay T between the locally generated and desired signal code sequences. This can be seen from the correlation diagram in Fig. 4. Due to the stepping nature of this scheme, the worst case correlation will occur for a delay of $\pm T_h/2n$. Assuming equally likely delays, the average loss from full correlation can be found. In the calculation of the parameter S_d , the desired signal energy must be multiplied by the factor

$$d \approx \frac{2n - 1}{2n} \quad (19)$$

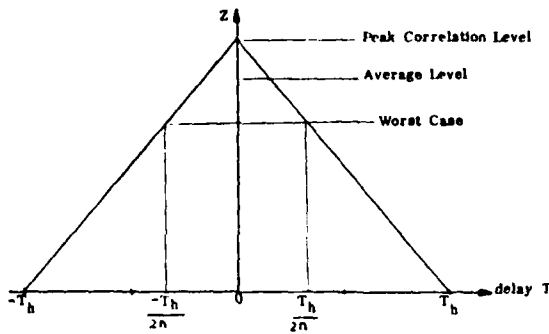


Fig. 4. Signal correlation diagram.

It should be noted that this scaling is used only for the stepped serial search scheme. This is because, in practical implementations of both the matched filter and two level schemes, the *largest* output of the passive correlator (in the interval where that output exceeds the threshold) would be used to determine the coarse code epoch, so there will be essentially no correlation loss.

Since each cell is examined for a period MT_h the search time for no false hits would be

$$T_s = (nN_c + 2)MT_h. \quad (20)$$

False hits tend to increase this time, but since P_{FH} is small the effect will not be significant.

Matched Filter Acquisition

The analysis is straightforward. False alarm and miss probabilities are given by (15) and (16). Since the search time is just the time taken to search the uncertainty region once,

$$T_s = (N_c + M)T_h. \quad (21)$$

Two-Level Acquisition

This scheme is analyzed in [4]. A false alarm occurs when the matched filter gives a false start signal which finds an idle correlator, which in turn generates the false synch indication. The probability of this occurring is

$$P_{FA} = P_{FA1}P_{FA2}(1 - B(c, g)) \quad (22)$$

where P_{FA1} and P_{FA2} are the false alarms conditions for the passive and active correlators, respectively, and $B(c, g)$ is the blocking probability for the queue of c active correlators given by the Erlang loss formula with $g = KP_{FA1}$.

A correct code acquisition epoch will be missed either when it is missed by the matched filter, or when it is correctly detected by the matched filter but is either missed by an active correlator or finds all active correlators engaged. The probability of this occurring is

$$P_M = P_{M1} + (1 - P_{M1})[B(c, g) + (1 - B(c, g))P_{M2}] \quad (23)$$

where P_{M1} and P_{M2} are the miss probabilities for the passive and active correlators, respectively.

The detection and false alarm probabilities for the sub-systems are provided by (15) and (16) with K substituted for M in the case of the active correlators

The search time is given by

$$T_s = (N_c + M + K)T_h. \quad (24)$$

VI. RESULTS

Starting with the serial search scheme, the basic performance curve is presented in Fig. 5 as the miss probability versus energy per chip to noise density ratio for various integration periods M (hops) in a benign environment. It is seen, for example, that a miss probability of about 10^{-3} is obtainable for, say, $E_c/N_0 = 5$ dB with M of 20. To take full advantage of this scheme, performance can be improved by increasing M without increasing hardware requirements but increasing search time. However, the accompanying set of results for an environment where half the received energy is in the scatter component and there are five interferers present shows that longer integration times are in fact necessary to produce miss probabilities of the same order.

The performance curves for a matched filter scheme are provided in Fig. 6. If the matched filter complexity M is equated to the stepped serial search integration time in hops, we find that the serial search scheme is slightly superior in performance. This is due to the fact that the serial search scheme considered requires three successive hits for a synch indication, whereas the matched filter relies on a single threshold exceedence.

With the two-level scheme it is meaningful to seek those thresholds L_1 and L_2 which minimize the miss probability, while the false alarm probability is held fixed. For each given parameter set consisting of M , K , c , E_c/N_0 , relative strength of scatter component b , and number of users N , the specified false alarm probability establishes L_2 for a fixed L_1 . Thus, each computation of miss probability is minimized over L_1 subject to the constraint of a given false alarm probability. Curves for the two-level scheme are presented in Fig. 7. Here schemes with one and three active correlators are compared. The active correlation period is 10 hops. It can be seen that an improvement in performance of 1-2 dB in the region of interest can be gained by using a bank of three active correlators as opposed to one. Comparison with the matched filter results show that, for similar performance, the two-level scheme with $K = 10$ requires a passive filter complexity only half of the matched filter on its own.

VII. CONCLUSIONS

The serial search scheme has been shown to provide good detection capabilities even in adverse environments and is relatively simple to implement. The cost is a long uncertainty region search time compared to that for the matched filter or two-level scheme. If this can be made small, such as in systems with reasonably fast hop rates and short codes, or if some rough code epoch can be maintained with clocks, then the acquisition times obtained can be acceptable, and the system is preferable in terms of detection reliability.

Matched filter detection requires a complex hardware

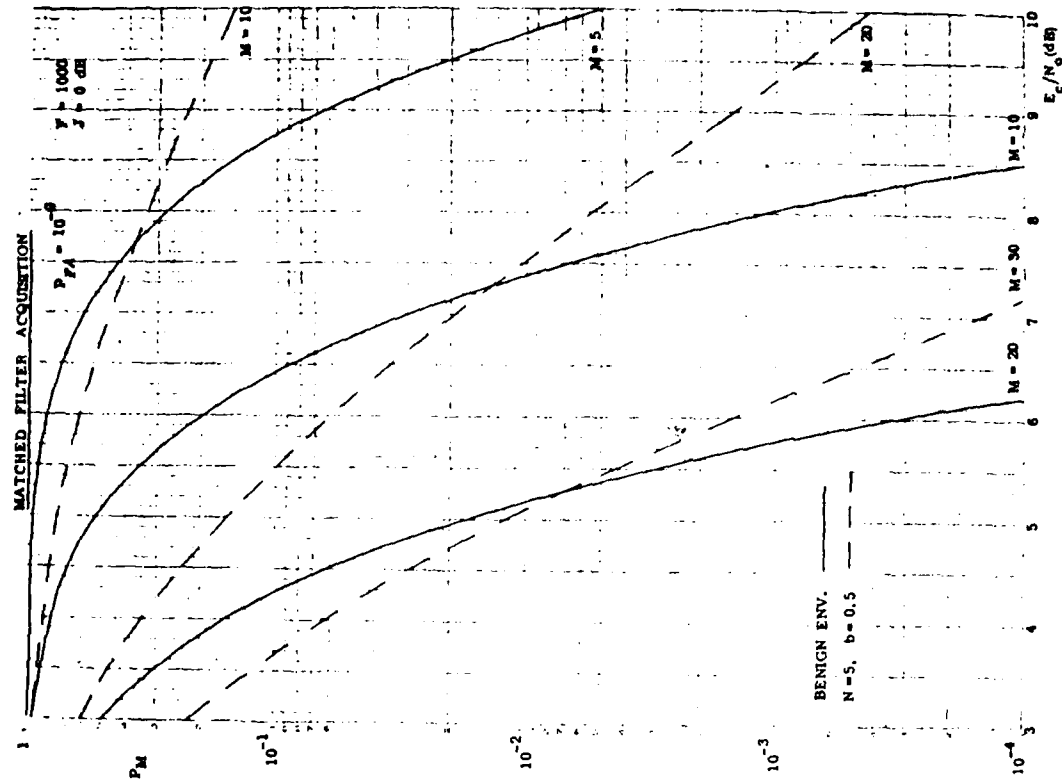


Fig. 6. Miss probabilities for matched filter acquisition.

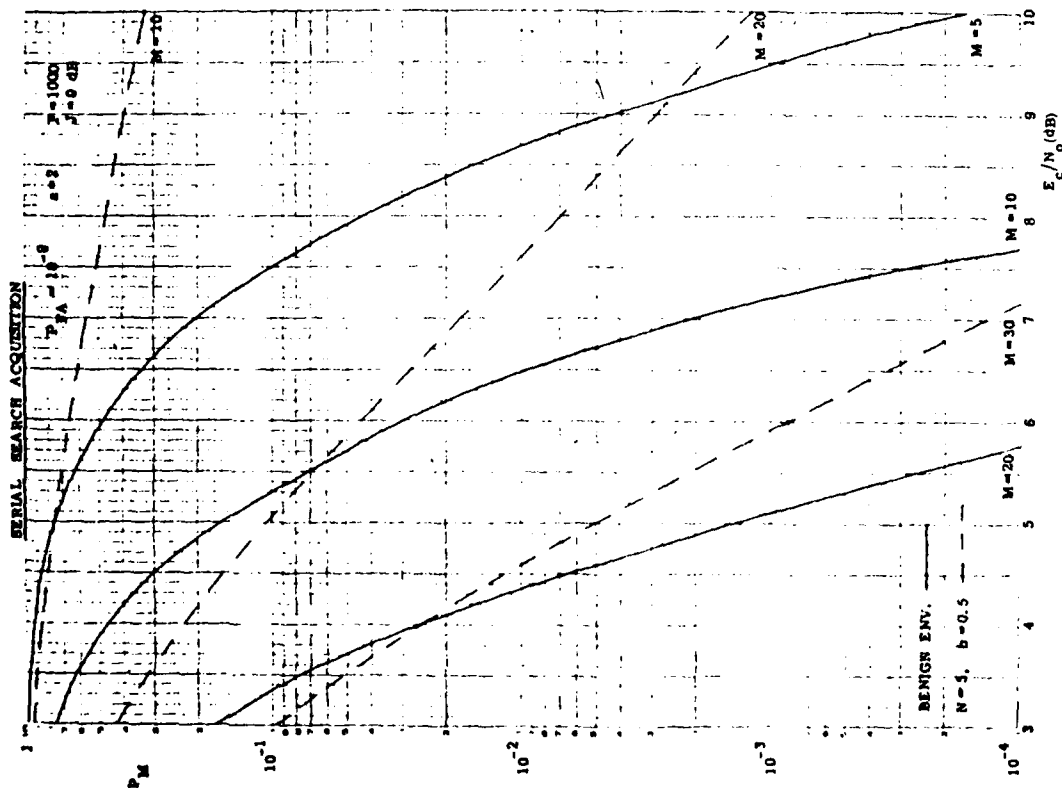


Fig. 5. Miss probabilities for serial search acquisition.

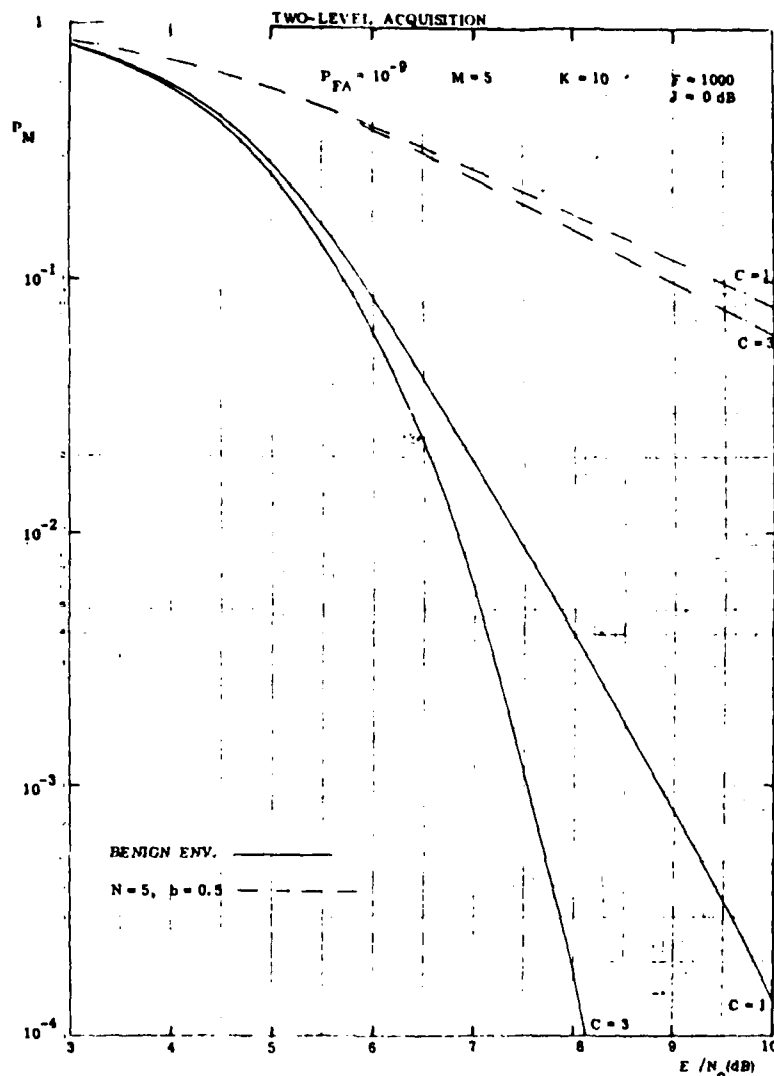


Fig. 7. Miss probabilities for two-level acquisition.

structure for similar performance. However, the matched filter searches the uncertainty region in real time, offering a clear advantage where it performs adequately. It is therefore preferable in systems operating in less adverse environments where simpler filters will provide the desired performance.

The two-level scheme combines the advantages of the rapid search time of the matched filter with the detection reliability of an active correlator. It would therefore find application in systems which do not rely on a time reference and which utilize long codes but require rapid and reliable acquisition in adverse environments. In situations with impaired signal detectability due to interference and/or scatter, the parallel bank of active correlators in this scheme provides additional versatility. Roughly, even while some correlators which may have been falsely engaged are operating, successful signal acquisition is still possible with an idle correlator. This parallelism is not available in the other schemes considered here. These advantages are at the expense of some-

what higher complexity, which, however, is not necessarily formidable.

REFERENCES

- [1] R. C. Dixon, *Spread Spectrum Systems*. New York: Wiley, 1976.
- [2] C. R. Cahn, "Spread spectrum applications and state-of-the-art equipments" (Agard Lecture Series No. 58).
- [3] P. Fieret *et al.*, "Applications of spread spectrum radio to wireless communications," presented at the Nat. Telecommun. Conf., Dec. 1980.
- [4] S. S. Rappaport and D. L. Schilling, "A two-level coarse code acquisition scheme for spread spectrum radio," *IEEE Trans. Commun.*, vol. COM-28, pp. 1734-1748, Sept. 1980.
- [5] A. D. Whalen, *Detection of Signals in Noise*. New York: Academic, 1971.
- [6] C. W. Helstrom, *Statistical Theory of Signal Detection*. London, England: Pergamon, 1968.
- [7] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.
- [8] P. M. Hopkins, "A unified analysis of pseudonoise synchronization by envelope correlation," *IEEE Trans. Commun.*, vol. COM-25, Aug. 1977.



Clive A. Putman was born in Durban, South Africa, on October 30, 1952. He received the B.Sc. Eng. degree from the University of Natal, South Africa, and the M.Sc. Eng. degree from the Cranfield Institute of Technology, Cranfield, England, in 1974 and 1977, respectively.

From 1975 to 1976 he was with Barlows Television Company, South Africa, and in 1978 he joined Marconi Instruments Ltd., England. Since 1979 he has been employed by Barlows Communications, South Africa, and is currently

working toward the Ph.D. degree at the City College of New York, New York, NY.



Stephen S. Rappaport (S'65-M'65-SM'76) received the B.E.E. degree from the Cooper Union, New York, NY, in 1960, the M.S.E.E. degree from the University of Southern California, Los Angeles, in 1962, and the Ph.D. degree in electrical engineering from New York University, New York, in 1965.

From 1960 to 1962 he was a member of the Technical Staff at Hughes Aircraft Company, where he was engaged in systems analysis for communications and radar. He was on the faculty at N.Y.U. from 1964 to 1965, and in 1965 joined Bell Laboratories, where he performed research and preliminary development of signal processing for radar and digital data communications. Since 1968 he has been on the Faculty of Electrical Engineering, State University of New York at Stony Brook, where he has served variously as Graduate Program Director and as Undergraduate Program Director. His present interests are in communications systems and techniques, multiple access, system modeling, queuing, communications traffic, and spread-spectrum techniques. A regular contributor to the technical literature, he has been a consultant to a number of industrial firms.

Dr. Rappaport is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi. He has been active in IEEE affairs. He was elected to the Advisory Council of the Communications Society for a two-year term which began January 1981, and he has served on the Communications Society Conference Board. In the past he served as Treasurer of the Long Island

Section, Chairman of the L. I. Communications Society Chapter, and Associate Editor for *COMMUNICATIONS MAGAZINE*, as well as Session Chairman, Organizer, and Technical Program Committee Member for some of the Society's major conferences. He is currently Associate Editor for Data Communications Systems for the *IEEE TRANSACTIONS ON COMMUNICATIONS*.



Donald L. Schilling (S'56-M'58-SM'69-F'75) was born in Brooklyn, NY, on June 11, 1935. He received the B.E.E. degree from the City College of New York, New York, NY, the M.S.E.E. degree from Columbia University, New York, and the Ph.D. degree in electrical engineering from the Polytechnic Institute of Brooklyn, Brooklyn, NY, in 1956, 1958, and 1962, respectively.

From 1956 to 1962 he was a Lecturer in the Department of Electrical Engineering, City College of New York, a Lecturer in the Department of Physics, Brooklyn College, Brooklyn, and a member of the Technical Staff of the Electronic Research Laboratories, Columbia University. In 1962 he was appointed Assistant Professor in the Department of Electrical Engineering, Polytechnic Institute of Brooklyn, and in 1966 he became an Associate Professor. He is currently with the Department of Electrical Engineering, City College of New York, and was appointed Distinguished Professor of Electrical Engineering through the award of the Herbert G. Kayser Chair in September 1980. He is also President of SCS Telecom, Inc. (formerly SCS Consulting Services), Sands Point, NY. In this position he supervises research and development in the military and commercial aspects of telecommunications, as well as directing an elaborate short course program in telecommunications, the computer, and management. He has coauthored four textbooks, *Electronic Circuits: Discrete and Integrated* (1968, 2nd ed. 1979), *Principles of Communication Systems* (1971), *Digital and Analog Systems, Circuits and Devices* (1973), and *Digital Integrated Electronics* (1977), and is currently writing a book on microcomputer applications. He has published more than 100 papers in the telecommunications field.

Dr. Schilling is a member of the Board of Directors of the IEEE. He was President of the IEEE Communications Society from 1979 to 1981 and was Director of Publications and Editor of the *IEEE TRANSACTIONS ON COMMUNICATIONS* from 1968 to 1978. He is a member of Sigma Xi and Eta Kappa Nu.

TRACKING OF FREQUENCY HOPPED SPREAD SPECTRUM SIGNALS IN ADVERSE ENVIRONMENTS

C.A. PUTMAN
Dept. of Elec. Engg.
City College of N.Y.
New York, N.Y. 10031

S.S. RAPPAFORT
Dept. of Elec. Engg.
State Univ. of N.Y.
Stony Brook, N.Y. 11794

D.L. SCHILLING
Dept. of Elec. Engg.
City College of N.Y.
New York, N.Y. 10031

ABSTRACT

Tracking requirements for a frequency hopped spread spectrum system with a view to application in mobile radio are investigated. Models characterizing Rician fading and multiple access or jamming environments are used to formulate signal detection probabilities. These determine the mean time to loss of lock. Optimized tracking in these environments is then considered.

1. INTRODUCTION

The frequency hopping (FH) waveform as a pseudo-random modulation technique has the attraction of ease of implementation and inherent frequency diversity [1]. A central feature of a FH system is the pseudorandom code generators at both the transmitter and receiver, capable of producing identical codes with proper synchronization. The pseudorandom code sequence is used to switch the carrier frequency via a frequency synthesizer and wideband mixer. When the synthesizer in the receiver is switched with the synchronized sequence the frequency hops on the received signal will be removed, leaving the original modulated signal.

The synchronization of FH spread spectrum signals can be broadly divided into two phases: acquisition and tracking. A typical system would first obtain coarse acquisition of the received pseudo-noise encoded signal [2], followed by fine acquisition by the tracking loop. At the end of the acquisition period loop parameters would be adjusted from acquisition to tracking settings and the system is said to be in lock. There are two possibilities at this point. Depending on the scheme used, either one or a series of false alarms may have caused the system to incorrectly enter the lock state, or the correct code epoch may have been detected causing the system to correctly enter the lock state. The requirements for the tracking phase are to maintain tracking of the correct code phase in the latter case, yet to quickly reject the false lock condition to allow resumed search for the correct code phase in the latter case. Although this paper focuses primarily on tracking of FH signals, much of the analysis applies directly or with little modification to other spread spectrum schemes such as direct sequence or hybrid systems. In-lock detection reliability is characterized by mean time to loss of lock when mean time to reject false lock is given.

2. THE TRACKING LOOP

The coarse acquisition schemes described in [2]

will provide acquisition of the received pseudo-random modulating sequence to within about half a code bit, or chip. Since finer acquisition is needed, and since reliance on clock stability to maintain accurate synchronization is generally not sufficient, even for short transmission push to talk operation, a tracking loop is usually employed.

The early-late gate tracking loop as used for range-gate tracking in radar systems [3], lends itself well to the frequency hopped signal application. This loop and its associated waveforms are shown in Figs. 1a,b. The gating waveform $g(t)$ from the VCO is alternately negative and positive with the minus to plus transitions coinciding with the chip to chip frequency hopping instants of the locally generated waveform $S_c(t)$. The error signal is generated by integrating the multiplied gating signal and envelope detector output $v(t)$ and is proportional to the delay T between the local and incoming code phases. The resulting discriminator characteristic is shown in Fig. 1c. The error voltage is used to advance or retard the local waveform into alignment.

Inherent in the loop operation will be some phase jitter which will increase with decreasing signal to noise ratio. The variance of this phase error is given in [3] for a single received pulse as

$$\sigma_T^2 = \frac{T_c T_p}{8E/N_0}$$

where T_c is the gate width, T_p is the pulse width and E/N_0 is the pulse energy to single-sided noise power spectral density ratio. For a frequency hopped system T_c is equal to the chip duration T_c , and assuming that the channel correlation bandwidth is greater than the hop rate, T_p is also equal to T_c . Hence

$$\sigma_T^2 = \frac{T_c^2}{8mE/N_0} \quad (1)$$

where m is the number of hops in the integration period and E_c is the energy per chip. This result also agrees with that in [4] with the loop bandwidth written as $B_L = 1/mT_c$.

3. IN-LOCK MONITORING

Once the system enters the lock state, it becomes necessary to monitor the tracking of the received code phase. This is achieved using the detector of Fig. 2a. The input waveform in the i th hop interval can be characterized as a narrowband waveform.

$$S_R(t) = \sqrt{2P_s} \cos[\omega_1(t - iT_c) + \theta_1] + g_i(t - iT_c) \quad (2)$$

where w_1 is determined by a known pseudorandom sequence, and $g_1(t)$ is a gaussian process with variance σ^2 . The envelope of this waveform is then a Rician distributed process, and since integration is over M hops, the output Z is the sum of squares of M independent Rician variates. For convenience we define a normalized variable Z_0 , such that $Z_0 = Z/\sigma^2$. Its first order probability density is then given by [5]

$$P_{Z_0}(z_0) = \frac{1}{2} (z_0/S) e^{-(z_0+S)/2} I_{M-1}(\sqrt{S}z_0)$$

where I_{M-1} is the Modified Bessel function of the first kind, order $M-1$, and the noncentrality parameter is

$$S = 2MP_s/\sigma^2. \quad (3)$$

If after MT_c seconds the output exceeds a threshold level V_T a synch detection is declared. The detection probability is then

$$P_D = P(Z > V_T) = \int_L^\infty P_{Z_0}(z_0) dz_0 \quad (4)$$

where $L = V_T/\sigma^2$.

This can be expressed in terms of the generalized Marcum Q-function as

$$P_D = Q_M(\sqrt{S}, \sqrt{L}) \quad (5)$$

where $Q_M(a, b) = \int_b^\infty x(x/a)^{M-1} e^{-(a^2+x^2)/2} I_{M-1}(ax) dx$.

Its complement is the miss probability expressed as

$$P_M = 1 - P_D = Q_M^c(\sqrt{S}, \sqrt{L}). \quad (6)$$

The false alarm probability for this system is the probability that the threshold is exceeded when the correct code sequence is not present. In the case of the benign environment considered thus far, the false alarm probability is

$$P_{FA} = Q_M(0, \sqrt{L}). \quad (7)$$

The tracking loop phase error will reduce the correlation between received and locally generated waveforms, thereby degrading in-lock detector performance, and must be included in the detection probability formulations. This is done by assuming that the phase error is approximately Gaussian distributed [3] so that 90 percent of the time the phase error is less than $3\sigma_T$. It can then be seen from the diagram of Fig. 2b that the correlation is reduced by a factor less than $3\sigma_T/T_c$ and hence the signal power at the detector output is scaled by a factor greater than

$$d = (1 - 3\sigma_T/T_c)^2$$

Hence the miss probability is given by (6) with, from (3),

$$S = \frac{2MdP_s}{B_{IF}N_0} \quad (8)$$

where B_{IF} is the system noise bandwidth.

Combining the tracking loop with a stepped serial search acquisition loop results in the circuit of Fig. 2c, with each loop providing fine and coarse VCO control respectively. Separate envelope detectors are used for tracking and signal detection since it is likely that their characteristics will differ. It can be noted that the in-lock detector is inherent in the same circuit with probable parameter changes. However, if matched filter acquisition were utilized, such circuitry would be additional.

5. CONTROL STRATEGY

Information received from the threshold detector, that is whether or not the threshold is being exceeded, is interpreted by a control system which decides whether or not to continue tracking or return to the search phase. This decision will be based on some form of control strategy [6]. In its simplest form, a single occurrence of the threshold not being exceeded, a miss, causes the return to the search phase. Alternatively a series of such occurrences may be required before this action takes place as shown in Figs. 3a,b. Here n is the number of additional in-lock states. The difference between the two strategies shown is that a threshold exceedence, or hit, either causes control to return to the first in-lock state or the previous one.

The control strategy can be represented by a finite Markov chain [7]. It is assumed that the probabilities of a hit or a miss are constants. The Markov chains for the strategies of Figs. 3a,b are shown in Figs. 4a,b respectively, where p is the probability of a hit and q is the miss probability. The first state has a reflecting wall, and the last state is an absorbing state representing rejection of the current code phase. The Markov chain can be described in terms of its transition matrix P whose element P_{ij} is the probability of transition from state i to state j . For the Markov chain of Fig. 4a, the $(n+2) \times (n+2)$ transition matrix is

$$P_a = \begin{bmatrix} p & q & 0 & \dots & 0 \\ p & 0 & q & & \\ 0 & p & 0 & & \\ \vdots & & & \ddots & \\ \vdots & & & & 0 & q & 0 \\ \vdots & & & & & p & 0 & q \\ 0 & \dots & \dots & 0 & 0 & 1 \end{bmatrix}$$

and for that of Fig. 4b,

$$P_b = \begin{bmatrix} p & q & 0 & \dots & 0 \\ p & 0 & q & & \\ p & 0 & 0 & & \\ \vdots & & & \ddots & \\ \vdots & & & & 0 & q & 0 \\ p & & & & & 0 & 0 & q \\ 0 & \dots & \dots & 0 & 0 & 1 \end{bmatrix}$$

respectively. In each case the matrix can be partitioned as

$$P = \begin{bmatrix} Q & R \\ 0 & 1 \end{bmatrix}$$

so that the $(n+1) \times (n+1)$ matrix Q describes the Markov chain without the absorbing state, R is a column vector of n zeros and a q , and 0 is a vector

of $n+1$ zeros.

The mean transition time from any state i to absorbing state $n+1$ is given by the i th element of the vector [7]

$$\underline{T} = (I - Q)^{-1} \underline{t} = N^{-1} \underline{t}$$

where I is an identity matrix and \underline{t} is a vector whose elements t_i are the times spent in state i . The mean time to loss of lock T_L is thus given by the first element of \underline{T} , the mean transition time from state 0 to state $n+1$. Using elementary matrix algebra,

$$T_L = \sum_{i=0}^n \frac{C_{i0} t_i}{\det N}$$

where C_{ij} is the cofactor of n_{ij} . If t_i , the threshold detector integration times, are a constant MT_C , then

$$T_L = MT_C \sum_{i=0}^n \frac{C_{i0}}{\det N} \quad (9)$$

This formula applies to any control strategy and is easily calculated when matrix algebra routines are available.

A slightly different approach yields closed form expressions for a given control strategy of n stages. The state transition diagrams can be transformed into signal flow graphs [8]. The strategies of Figs. 3a,b can then also be represented by the signal flow graphs of Figs. 5a,b where the variable t represents the unit path length, in this case the fixed integration time. The mean path length from node 0 to node $n+1$ is then the mean time to loss of lock. Since the mean path length is just the sum of all path lengths times their probability it can be obtained from the graph transfer function $G_n(t)$ by differentiating with respect to t and setting $t=1$ [8]:

$$T_L = MT_C \left. \frac{dG_n(t)}{dt} \right|_{t=1} \quad (10)$$

The graph transfer function for Fig. 5b is obtained by inspection using Mason's formula [9] as

$$G_n(t) = \frac{q^{n+1} t^{n+1}}{1 - p t \sum_{i=0}^n q^i t^i}$$

Applying equation (10) and using $p = 1-q$ it is found that

$$T_L = MT_C \frac{1-q^{n+1}}{(1-q)q^{n+1}}$$

Note that as the miss probability approaches unity we find

$$\lim_{q \rightarrow 1} T_L = M(n+1) \quad (11)$$

This result will be used as a basis for comparison of various control strategies. A general closed form expression for T_L for the graph of Fig. 5a has not been found, but the same technique yields an expression for each value of n . In each case

equation (11) applies.

This is in order to determine the mean time to loss of lock given that the correct code epoch has been detected, or the mean hold-in time T_H , we substitute $1-q$ for q in the expression obtained for T_L for the control strategy used. Similarly, substituting $1-1/F$ for q yields the mean time to loss of lock given incorrect acquisition, or mean time for false lock T_{FL} .

5. ENVIRONMENT MODELING Fading Channel Model

We consider a Rician fading channel so that the received signal with which synchronization is desired is given during the i th hop interval by

$$S_R(t) = \sqrt{2a_i P_s} \cos(\omega_i t - \theta_i) + \sqrt{2b_i P_s} \cos(\omega_i t - \theta_j) + n_w(t) \quad (11)$$

consisting of specular, scatter and noise components respectively. The scatter and noise components are Gaussian and independent and the scatter component is signal dependent. Equation (11) defines the fading model and for $a_i=0$ is identical to that given in [10].

For comparative analysis we equate the sum of powers in the specular and scatter components to a constant, P_s , representing the average received signal power over many hops. In particular we take

$$\bar{a} + b = 1$$

$$\text{where } \bar{a} = \frac{1}{M} \sum_{i=1}^M a_i \text{ and } b = \sum_{j=1}^J c_j^2$$

Here \bar{a} is summed over many hops and b is summed over many multipaths. The quantities \bar{a} and b are indicative of the relative strength of the specular and scatter components respectively. It is assumed that these parameters are fixed over the acquisition period, although they may have long term slow variations.

Following a derivation similar to that in [10], the above signal can be expressed in the form of (2). The noncentrality parameter S in the equation for the miss probability (6) then becomes

$$S = 2M\bar{a}bP_s/\sigma^2 \quad (13)$$

$$\text{where } \sigma^2 = N_0 B_{IF} + bP_s$$

Multiple access model

We now consider the interference generated by other users who are hopping over the set of F frequency slots. The probability that no other user causes interference during a given hop when N other users are active is $(1-1/F)^N$. The probability that at least one other user causes interference is

$$p = 1 - (1-1/F)^N \quad (14)$$

Since we are integrating over M hops, the probability that J slots have at least one user present, is

$$b(J,M,p) = \binom{M}{J} p^J (1-p)^{M-J} \quad (15)$$

Jamming Model

In application the receiver could provide for limiting of received signal strength above that of the desired signal. This mitigates the effects of CW jamming. An optimized jammer would be forced

to jam a random selection of frequencies. When the number of frequencies available F is large, the above model is reasonable for random hop or comb jamming, if N is taken as the number of interferers or interfering tones respectively.

The Combined Model

To quantify performance for a fading environment with interferers present it is analytically convenient to assume that the power of interfering signals lies in the specular components and hopped interfering signals are assumed to be synchronous with the desired signal. In addition, all interferers are assumed to be of equal strength, and where interferers and desired signal are present in the same slot their powers are assumed to be additive.

The false alarm probability given J slots are jammed is then

$$P(\text{false alarm} | J) = Q_M(\sqrt{S_J}, \sqrt{L}) \quad (16)$$

with

$$S_J = J2JP_s / N_o B_{IF} = J2JE_c / N_o$$

where J is the jammer power (or limited signal power) to desired signal power ratio and we assume that $B_{IF} = 1/T_c$. The overall false alarm probability is then determined by multiplying by the probability of J slots being jammed and summing over J ,

$$P_{FA} = \sum_{J=0}^M b(J, M, p) Q_M(\sqrt{S_J}, \sqrt{L}) \quad (17)$$

Similarly the miss probability is

$$P_M = \sum_{J=0}^M b(J, M, p) Q_M^c(\sqrt{S_d}, \sqrt{L'}) \quad (18)$$

with

$$S_d = \frac{[J(J+a) + (M-J)a]2P_s}{N_o B_{IF} + bP_s} = \frac{[JJ + Ma]2E_c / N_o}{1 + bE_c / N_o}$$

and

$$L' = L / (1 + bE_c / N_o).$$

6. RESULTS

The performance parameter of interest is the mean hold-in time T_H given the mean time for false lock T_{FL} . For graphical convenience we generate the logarithm of these times normalized to the hop period T_c . Thus given T_{FL} , M , m , and n , P_{FA} is specified so that for any set of parameters E_c/N_o , relative strength of scatter component b , and number of interferers N , the detector threshold level and hence P_M and finally T_H can be calculated.

Consider a tracking system employing the strategy of Fig. 5a. The basic performance curve is presented in Fig. 6 as the log of normalized mean hold-in time vs energy per chip to noise density ratio for varying number of control strategy states n . For the loop parameters M and m chosen, the need for a control strategy with one or more additional states is evident. For example, if the mean hold in time is required to be greater than an hour at an E_c/N_o of 6dB with a hop rate of 10KHz and a mean time to reject false lock of 10mS, then a value of zero for n (i.e., a single miss causes immediate return to the search phase) is

sufficient in a benign environment. However, when the environment consists of a fading channel where half the received energy is in the scatter component and there are 50 interferers present, one or two additional in-lock states are needed.

Minimum tracking system requirements for a specified mean hold-in time in a given environment are determined by the specified mean time to reject false lock. To see this consider the curves of Fig. 7, log of normalized mean hold-in time vs log of normalized false lock time, which show the trade-off in performance for reduced mean false lock time. With reference to the results for an environment where half the received energy is in the scatter components and there are 50 interferers present, we see that for the parameters chosen, if T_{FL} is required to be less than 10ms then two additional in-lock states should be used.

The curves of Figs. 8 and 9 show the effect of the relative strength of scatter component and number of interferers respectively on the mean hold-in time. The effect of the fading environment is seen to be more severe than that of the multiple-user or jamming environment, and control strategy requirements can be determined from these results.

Similar results were obtained for a system using the control strategy of Fig. 3b, and were found to have little significant difference from the above results for the values of n considered. Either strategy would thus suffice.

Finally, there were also found to be little significant difference in the performance obtained from a system with $n=0$ and $M=30$ compared to a system with $n=2$ and $M=10$ (application of equation (11) yields a normalized mean time to loss of lock of 30 in each case). This implies that adequate performance can always be obtained simply by increasing the in-lock detection integration time, provided the constraint $T_{FL}/T_c \gg M$ is met. However, the analytical results predict average performance and do not reflect the disastrous effect of events such as deep fades which are often prevalent in a mobile environment. The in-lock states and shorter integration times would be preferable since a single miss does not indicate loss of lock.

7. CONCLUSIONS

A system for the tracking of frequency hopped signals has been evaluated in terms of mean hold in time for given mean time to reject false lock. It was found that tracking system requirements for a benign environment are greatly extended by the Rician fading environment, and only to a limited degree by a multiple access or jamming environment. System design should therefore utilize the predicted results for an expected environment.

In the case of a benign environment, performance of the tracking system in terms of mean hold-in time can always be improved by increasing the in-lock detector integration time, providing that the mean time to reject false lock is acceptable. However, a search lock strategy with one or more additional in-lock states gives similar performance for the same mean time to reject false lock and is preferable in the case of the fading environment.

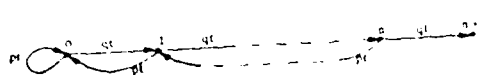
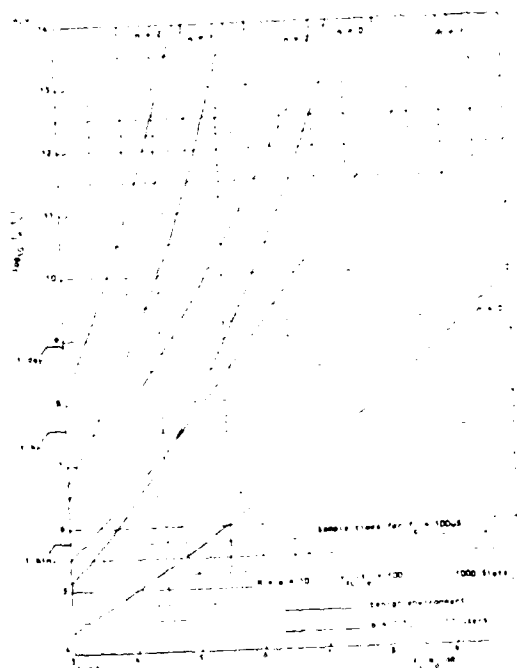
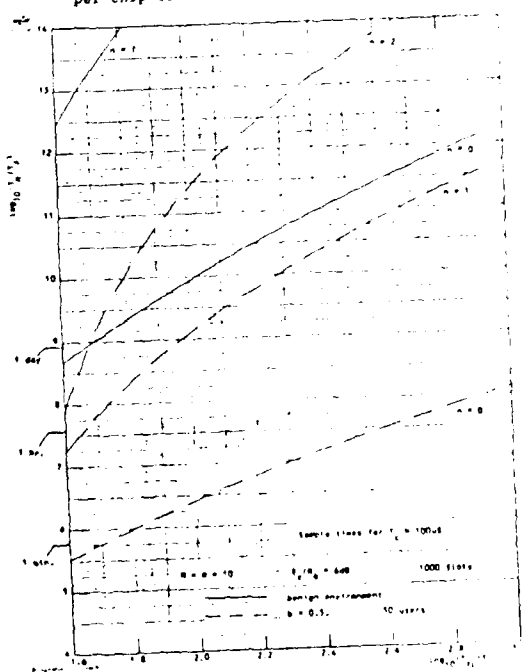


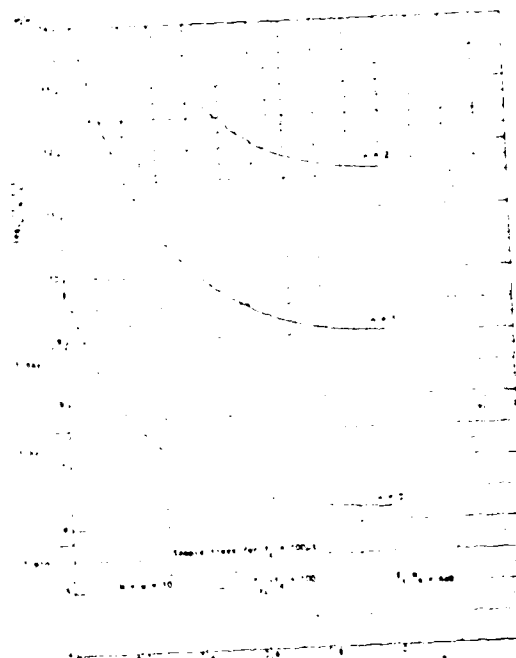
Fig. 3a, b. Signal flow graphs for Fig. 3a, b, resp.



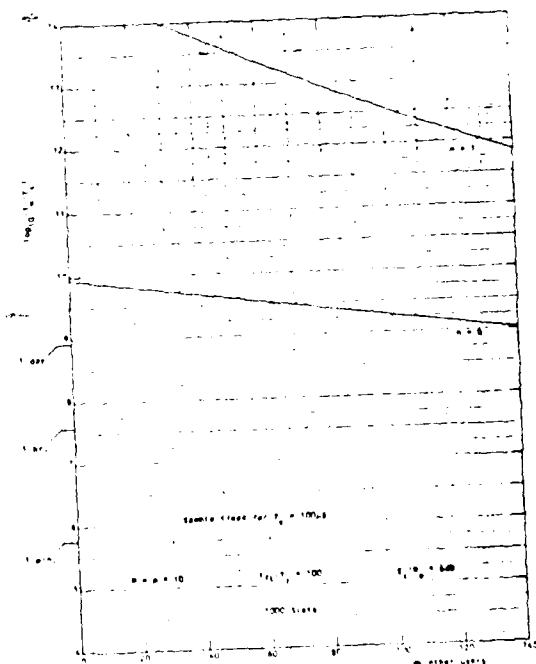
6. Log of normalized mean hold-in time vs energy per chip to noise density ratio.



7. Log of normalized mean hold-in time vs log of normalized mean time to reject false lock.



8. Log of normalized mean hold-in time vs relative strength of scatter component.



9. Log of normalized mean hold-in time vs number of other users.

Coherent optical production of the Hough transform

George Eichmann and B. Z. Dong

The Hough transform (HT) is an efficient shape detector that maps straight lines into a two-parameter feature space. Recently it has been pointed out that the forward Radon transform (FRT), well known from the theory of computed tomography, and the HT are equivalent for binary images. In this paper, analog coherent optical implementation of the FRT is discussed. The FRT will not only be of use in implementing the HT shape descriptors but also act as a coherent optical preprocessor for the implementation of multidimensional convolution, correlation, and spectral analysis using 1-D acoustooptical signal processing devices. Several different coherent optical FRT architectures are presented. Experimental results using conventional coherent Fourier transform configuration are given. The relationship between the coherent optical implementation of the FRT and the inverse Radon transform, an important tool in computed tomography, is also detailed.

I. Introduction

The Hough transform¹ for detecting straight lines has been discussed in the context of image pattern analysis.²⁻⁴ The Hough transformation maps a point in the x - y plane into a sine curve in the (p, θ) transform plane using the formula $p = x \cos \theta + y \sin \theta$. A straight line in the x - y plane is mapped into a point in the transform plane. Every picture edge element, which may be considered as a unit length line segment, is mapped into a point in the transform plane. All collinear edge elements are mapped into the same point in the transform plane. The total number of unit length collinear line segments form a histogram at a point in the transform plane. A single long or many short collinear line segments can yield the same histogram value. Images that contain edge elements only, the so-called shape function, can always be decomposed as a set of approximate line segments. The Hough transform maps these line segments into a smaller feature space to form the Hough transform shape descriptors.

Recently it has been pointed out by Deans⁵ that the Hough transform for a binary image is equivalent to the forward Radon transform⁶ (FRT) of this image, a transform that has been known for a half-century. In fact the generalized versions of the Hough transform⁷ that detect other shapes such as rectangles, ellipses, or parabolas can be shown to be equivalent to a generalized version of the FRT. In this case, the generalized FRT maps the 2-D shape function into a multidimensional feature space. The Radon transform theory has not only applications in pattern recognition but also in nuclear medicine, imaging by nuclear magnetic resonance, determination of the electron-momentum distribution in solids, scattering theory, and plasma diagnostics.⁸⁻¹⁰ While digital methods for both construc-

tion and reconstruction of the Radon transform are available,¹¹⁻¹⁴ optical analog reconstruction techniques have also been discussed.¹⁵⁻¹⁹ A recent survey of both coherent and incoherent optical reconstruction techniques is available.²⁰ More recently, Barrett²¹ has discussed use of incoherent optical FRT processors for performing multidimensional convolution, correlation, and spectral analysis using a series of 1-D optical Radon transform processors. In this paper, coherent optical implementations of the FRT are discussed. The purpose of the optical processors is to be used as a bandwidth compressor of optical shape functions as well as in the use of 1-D coherent optical signal processing of multidimensional signals.

II. Radon Transform

The FRT can be defined in a number of spaces and in an arbitrary number of dimensions.²² If $f(x, y)$ is the 2-D input distribution, the FRT of this function is defined as, using vector notation,

$$\hat{f}(p, \theta) = \int_{-\infty}^{\infty} f(\mathbf{x}) \delta(p - \mathbf{x} \cdot \boldsymbol{\xi}) d\mathbf{x}, \quad (1)$$

where

$$\mathbf{x} = x\mathbf{a}_1 + y\mathbf{a}_2, \quad (2)$$

$$\boldsymbol{\xi} = \cos \theta \mathbf{a}_1 + \sin \theta \mathbf{a}_2 = \xi_1 \mathbf{a}_1 + \xi_2 \mathbf{a}_2, \quad (3)$$

The transform parameters (p, θ) can be considered as defined on a rectangular grid in the (p, θ) plane. A point (p_0, θ_0) in the transform plane corresponds to a line in the x - y plane where p_0 is the shortest distance between a point on the line and the origin, and θ_0 is the angle formed by this shortest chord to the line measured from the positive x axis (Fig. 1). Conversely, lines passing through the point (x_0, y_0) in the x - y plane correspond to the sinusoidal curve

$$p = x_0 \cos \theta + y_0 \sin \theta \quad (4)$$

in the (p, θ) plane. These properties are the basis of the line detection capability of the FRT. While the present discussion deals with line detection in noiseless images, an analysis of curve detection in noisy images is avail-

The authors are with City College of the City University of New York, Department of Electrical Engineering, New York, New York 10031.

Received 7 September 1982.

0003-6935/83/060830-05\$01.00/0.

© 1983 Optical Society of America.

able.^{2,12,14} Some of the elementary properties of the FRT are now described.

Translation of the input transparency by a distance \mathbf{x}_0 , i.e.,

$$f_{x_0}(\mathbf{x}) = f(\mathbf{x} + \mathbf{x}_0), \quad (5)$$

yields a FRT

$$f_{x_0}(p, \theta) = f(p + \xi \cdot \mathbf{x}_0, \theta) \quad (6)$$

Taking the directional derivative of the input,

$$f_2(\mathbf{x}) = \mathbf{a} \cdot \nabla f(\mathbf{x}), \quad (7)$$

where \mathbf{a} is a constant vector, yields as its FRT

$$f_2(p, \theta) = (\mathbf{a} \cdot \xi) \partial f / \partial p, \quad (8)$$

while the linear multiplication of the input

$$f_3(\mathbf{x}) = (\mathbf{a} \cdot \mathbf{x}) f(\mathbf{x}) \quad (9)$$

yields a directional derivative in the angle FRT plane:

$$\partial f_3 / \partial p = -(\mathbf{a} \cdot \nabla_\theta)^2 f \quad (10)$$

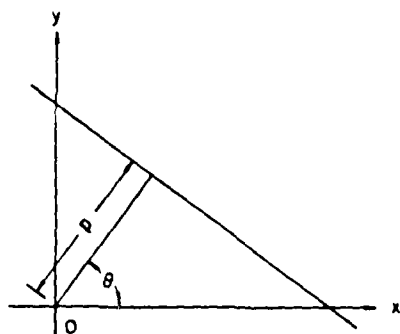


Fig. 1. Cartesian input and radon transform plane coordinate geometries. The shortest distance to the straight line is p , while θ is the angle formed by the shortest distance cord and the x axis.

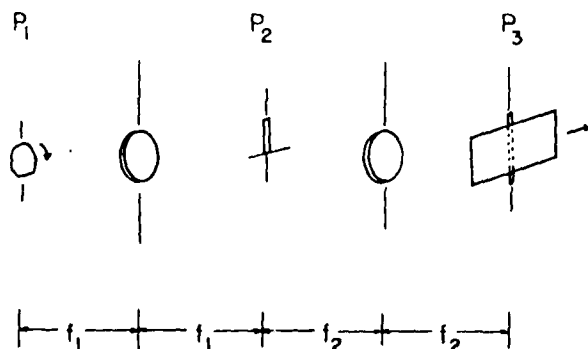


Fig. 2. Coherent Fourier optical lens configuration used to record the FRT of a 2-D input. The input is located in plane P_1 . A vertical slit is placed in the Fourier transform plane P_2 , while the output FRT is recorded in the inverse Fourier transform plane P_3 . For a fixed angle of the input in the plane P_1 , a Cartesian coordinate slice, with FRT coordinates as a Cartesian grid, is generated at the output plane P_3 . Rotating the input in plane P_1 by an incremental angle while simultaneously linearly translating the output storage material in the output plane P_3 , a new slice of the FRT picture is generated.

Finally the FRT of the 2-D convolution of two inputs $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ in the space domain

$$f(\mathbf{x}) = \int_{-\infty}^{\infty} f_1(\mathbf{y}) f_2(\mathbf{x} - \mathbf{y}) d\mathbf{y} \quad (11)$$

is

$$f(p, \theta) = \int_{-\infty}^{\infty} f_1(\theta, t) f_2(\theta, p - t) dt, \quad (12)$$

a 1-D convolution of the FRT of the inputs in the Radon transform (RT) domain. This property of the RT, namely, that it maps 2-D into 1-D convolution, can be extended into higher dimensions. Also this property allows reduction of multidimensional correlation, convolution, and spectral analysis to be performed using 1-D acoustooptical signal processing architectures.

III. Optical Methods of Generating the Hough Transform

The starting point for the coherent optical generation of the FRT is the direct implementation of the integral

$$f(p, \theta) = \iint_{-\infty}^{\infty} f(x, y) \delta(p - x \cos \theta - y \sin \theta) dx dy, \quad (1)$$

where $f(x, y)$ is the input and $f(p, \theta)$ is the output with a delta function as the space-variant impulse response kernel function. A particular way to implement this as well as other similar kernel functions is to combine suitable linear and/or circular motions that will map the space-variant kernel into a space-invariant kernel. If the coordinate axes (x, y) are rotated by an angle θ , the new coordinate axes (u, v) are

$$u = x \cos \theta + y \sin \theta \quad v = -x \sin \theta + y \cos \theta. \quad (13)$$

In the rotated coordinates the input transparency is $f^-(u, v)$. Thus by rotating the input, the output FRT can be expressed in terms of a space-invariant kernel operation

$$f(p, \theta) = \iint_{-\infty}^{\infty} f^-(u, v) \delta(p - u) du dv. \quad (14)$$

This integration can be performed either in the space or in the spatial frequency domain. In the space domain, for example, one could electronically integrate the rotated function along the v axis and then use a 2-D coherent optical processor to implement the second integration. This integration can readily be performed using an astigmatic processor.²⁵ Or one can use a 2-D multiplexed holographic processor to implement directly the space-variant impulse response function of Eq. (1). Alternatively, since the kernel of Eq. (14) is space-invariant, the filtering operation can be performed in the spatial frequency plane. Letting the Fourier transform of the rotated input transparency be $F^-(z, w)$, the FRT is

$$f(p, \theta) = \iint_{-\infty}^{\infty} F^-(z, w) H^-(z, w; p, \theta) dz dw, \quad (15)$$

where the filter function H^- is

$$H^-(z, w; p, \theta) = (1/2\pi) \delta(w) \exp(-jz p). \quad (16)$$

This filter can be implemented using a slit and linear phase-shifting wedge. An experimental implementa-

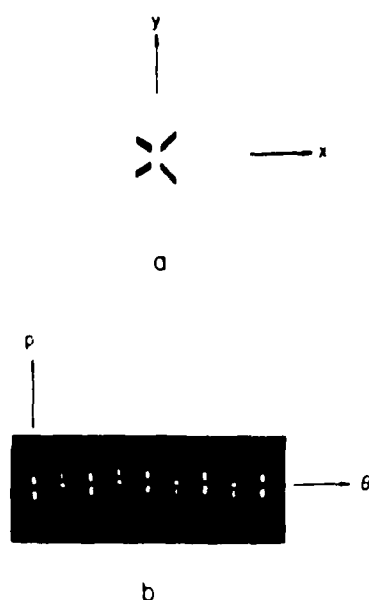


Fig. 3. Experimental results generated using the optical setup described in Fig. 2. In (a) the input while in (b) the FRT slices of (a) are shown. The vertical axis in (b) is p , while the horizontal is the θ axis. Here the angle is measured from the Cartesian y axis.

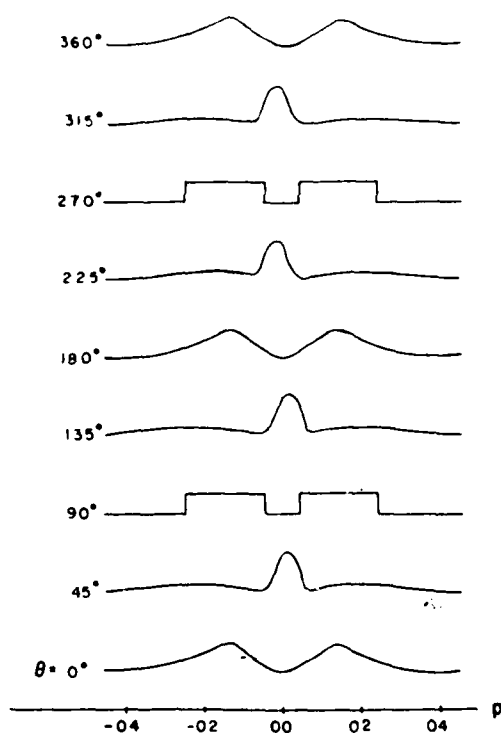


Fig. 4. Computer generated FRT slices for the input of Fig. 3(a).

tion of this scheme using a coherent Fourier optical processor is now presented.

In Fig. 2 a coherent Fourier optical lens configuration is used to record the FRT slices. The input transparency is located in the plane P_1 . A narrow vertical slit prescribed in Eq. (16) is placed in the transform plane

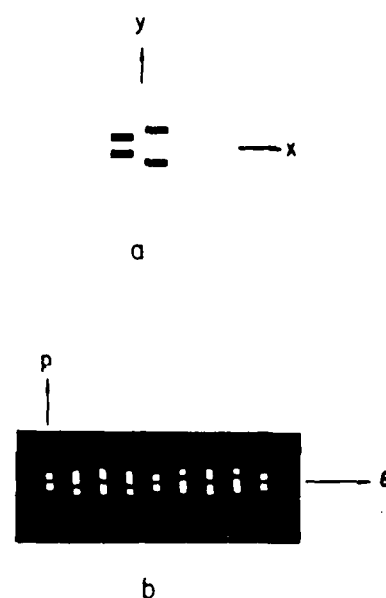


Fig. 5. Experimental results generated using the optical arrangement shown in Fig. 2. In (a) the straight lines input, while in (b) the FRT output slices of the input of (a) are shown.

P_2 . The 1-D linear phase shift is implemented using a second Fourier transform lens together with a 1-D slit in the output plane P_3 . In the film plane P_3 , for a given angle θ , a slice of the FRT plane is recorded. By rotating the input in plane P_1 , a new angle is selected. The new angle is recorded by linearly translating the film to a new Cartesian grid position in the FRT plane. In this fashion, the whole FRT output plane is filled sequentially. In Fig. 3(a), a particular input containing both horizontal and angled lines is shown. The resulting FRT slices are depicted in Fig. 3(b). The vertical is p , while the horizontal is the θ axis. The FRT slices are spaced 45° apart with zero at the left and 360° at the right. In Fig. 4, results of the computer generated slices for the input of Fig. 3 are shown. Note: to be consistent with the experimental arrangement, the angle θ is measured from the vertical axis. While the results are sensitive to the choice of the origin [see Eq. (6)], these results do match the experimentally obtained results of Fig. 3(b). In Fig. 5, a second input with the corresponding output FRT slices are shown. Here, as in the previous figure, the straight lines are thickened for the FRT analog computer. While the ideal HT descriptors deal with idealized line segments, practical boundaries are always filled for better detection. Thick boundaries are also necessary for the optical analog computer to improve the detection SNR. In Fig. 6, computer generated FRT slices of the input of Fig. 5(a) are shown. With the identical caveat mentioned previously, these results match the experimental results of Fig. 5(a).

Various other expressions for generating representations of the FRT can be derived by using different representations of the Dirac delta function. These

representations can lead to new optical implementation methods for the analog evaluation of the FRT operation. The kernel of the FRT, a space variant impulse response function, is a function of four variables. One may isolate any particular variable and derive new representations for the FRT. Consider a change from Cartesian (x, y) to cylindrical coordinates (r, ϕ) , where

$$x = r \cos \phi \quad y = r \sin \phi. \quad (17)$$

In cylindrical coordinates the FRT kernel function is

$$h(r, \phi; p, \theta) = \delta[p - r \cos(\theta - \phi)]. \quad (18)$$

Using elementary properties of the delta function, Eq. (18) can be expressed in terms of the radial variable as

$$h(r, \phi; p, \theta) = \delta[r - p/\cos(\theta - \phi)]/|\cos(\theta - \phi)| \quad (19)$$

or in terms of the angular variable as

$$h(r, \phi; p, \theta) = \delta[\phi - \theta + \cos^{-1}(p/r)]/r \sqrt{1 - (p/r)^2}. \quad (20)$$

These representations of the delta functions are now used to derive new expressions for the FRT.

Let the input function $f(r, \phi)$ be expressed as an angular Fourier series

$$f(r, \phi) = \sum_{n=-\infty}^{\infty} f_n(r) \exp(-jn\phi), \quad (21)$$

where

$$f_n(r) = \frac{1}{2\pi} \int_0^{2\pi} f(r, \phi) \exp(jn\phi) d\phi. \quad (22)$$

Substituting Eqs. (20) and (21) into Eq. (1) we have

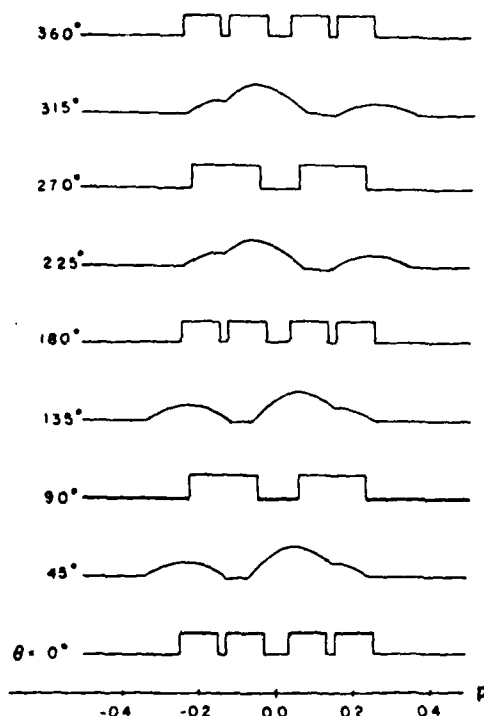


Fig. 6. Computer generated FRT slices for the input of Fig. 5(a).

$$\hat{f}(p, \theta) = \sum_{n=-\infty}^{\infty} \exp(-jn\theta) \hat{f}_n(p), \quad (23)$$

where

$$\hat{f}_n = \int_0^{\infty} f_n(r) [1 - (p/r)^2]^{-1/2} \exp[-jn \cos^{-1}(p/r)] dr. \quad (24)$$

The FRT is expressed as periodic Fourier series in the angular variable and a forward circular transform in the radial direction. The inverse circular transform has been discussed in the context of computed tomography and in terms of coherent optical implementation by Hansen and Goodman.¹⁶ The forward and the inverse circular transform kernels are quite similar, and, therefore, both of these transforms can be performed with the same optical configuration.

We now use the second radial expression of the kernel function of Eq. (19) to derive a different representation of the FRT. Let the input transparency be represented as

$$f(r, \phi) = \frac{1}{2\pi} \int_{Br} G(u, \phi) \exp(ur) du, \quad (25)$$

where

$$G(u, \phi) = \int_0^{\infty} f(r, \phi) \exp(-ur) dr \quad (26)$$

is the Laplace transform of the input transparency in the radial direction and Br stands for the Bromwich path necessary for evaluation of the inverse Laplace transform. Substituting Eqs. (25) and (19) into Eq. (1) we have a new representation:

$$\hat{f}(p, \theta) = \frac{1}{2\pi} \int_{Br} G^-(u; p, \theta) du, \quad (27)$$

where

$$G^-(u; p, \theta) = p \int_{-1}^1 (1 - u^2)^{1/2} / u G\{u, \cos^{-1}(1/u) + \theta\} \times \exp(-pu \cos \theta) du. \quad (28)$$

Here again the 2-D integration has been decomposed into two 1-D integrations. The optical evaluation of the Laplace transform operation has been discussed.²⁵ The second integration indicated by Eq. (28) can be performed by an astigmatic processor.²⁶

Another method of optically performing the FRT is based on a coordinate distortion technique. Coordinate distortion in optics has been discussed,^{27,28} and in particular use of coherent optical IRT has been described by Hofer.¹⁷ The method is based on a 2-D Cartesian form of the kernel function (1). Consider the following expansion of the delta function kernel:

$$\delta(p - x \cos \theta - y \sin \theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp[-ju(p - x \cos \theta - y \sin \theta)] du. \quad (29)$$

Using this representation of the delta function we can represent the FRT as

$$\hat{f}(p, \theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(u \cos \theta, u \sin \theta) \exp(-ju p) du, \quad (30)$$

where

$$F(x \cos \theta, y \sin \theta) = \iint_{-\infty}^{\infty} f(x, y) \exp[ju(x \cos \theta + y \sin \theta)] dx dy \quad (31)$$

Here the 2-D Fourier transformed input is used as a building block to perform the FRT. The 2-D Fourier transform is performed by a conventional coherent Fourier optic processor. By performing a 1-D Fourier transform in the radial and imaging in the θ direction on the 2-D Fourier transformed input, the FRT can be formed. The coordinate distortion using an astigmatic optical processor must be used to perform the radial to Cartesian coordinate conversion followed by a standard astigmatic optical processor that images in one and takes the Fourier transform in the other direction.

Finally another delta function representation can be useful in the case where both the input and its FRT are limited to a finite region in both the input and FRT plane. In this case, both the input and its FRT can be considered a periodic function. Consider the following resolution of the delta²⁹ function:

$$\delta(x - x') = \sum_{n=0}^{\infty} A_n \cos n\pi x/L \cos n\pi x'/L, \quad (32)$$

where

$$A_n = 1/L, n = 0 \text{ and } 2/L, n \neq 0. \quad (33)$$

Substituting Eq. (32) into Eq. (1) we have a periodic function description in the p direction of the FRT:

$$f(p, \theta) = \sum_{n=0}^{\infty} A_n \cos n\pi p/L g_n(\theta), \quad (34)$$

where

$$g_n(\theta) = \iint_{-\infty}^{\infty} f(x, y) \cos n\pi/L [x \cos \theta + y \sin \theta] dx dy \quad (35)$$

is the 2-D cosine transform of the input. The 2-D cosine transform can be generated using the conventional 2-D coherent optical processor.³⁰

IV. Summary and Conclusions

Analog coherent optical generation of the HT shape descriptors has been discussed. HT shape descriptors are one of the most efficient boundary descriptors in pattern recognition. A number of possible analog implementations have been described. The optical implementations are based on the various representations of the Dirac delta (identity) function, which is the space-variant impulse response of the HT kernel. Since the HT and the FRT are identical for binary images, the above implementations are identical to the optical generation of the FRT. Because the FRT of multidimensional convolution in the space domain is a 1-D convolution in the Radon transform domain, multidimensional coherent FRT processors can act as optical preprocessors for coherent 1-D acoustooptic signal processing architectures. Experimental results using 2-D coherent Fourier optical transform lens configuration together with rotational and linear motion have been presented. A good match between computer generated and experimental results is shown. A number of other coherent optical FRT architectures are proposed. Some of these architectures are similar to

those proposed for evaluation of the inverse Radon transform known in computed tomography.

This work was supported in part under a grant from the Air Force Office of Scientific Research AFOSR 81-0169 and a contract from the Rome Air Development Center F19628-80-C-0095.

References

1. P. V. C. Hough, "Methods and Means for Recognizing Complex Patterns," U.S. Patent 3,069,654 (1962).
2. W. K. Pratt, *Digital Image Processing* (Wiley-Interscience, New York, 1978).
3. S. D. Shapiro and A. Lannino, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-1**, 310 (1979).
4. R. O. Duda and P. E. Hart, *Commun. ACM* **15**, 11 (1972).
5. S. R. Deans, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-2**, 185 (1981).
6. J. Radon, *Ber. Sachs. Akad. Wiss. Leipzig* **69**, 262 (1917).
7. K. R. Sloan, Jr., *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-4**, 87 (1982).
8. R. Gordon, R. Bender, and G. T. Herman, *J. Theor. Biol.* **29**, 471 (1970).
9. W. M. Boerner, C. M. Ho, and B. Y. Foo, *IEEE Trans. Antennas Propag.* **AP-29**, 336 (1981).
10. R. C. Chase, F. H. Segum, M. Getasimenko, and R. Petrasco, *Proc. Soc. Photo-Opt. Instrum. Eng.* **231**, 265 (1980).
11. R. M. Mersereau and A. V. Oppenheim, *Proc. IEEE* **62**, 1319 (1974).
12. *Technical Digest, Topical Meeting on Imaging Processing for 2-D and 3-D Reconstructions from Projections* (Optical Society of America, Washington, D.C., 1975).
13. Z. H. Cho, Ed., Special issue on Physical and Computational Aspects of 3-D Image Reconstruction, *IEEE Trans. Nucl. Sci.* **NS-21** (June 1971).
14. B. K. Gilbert, A. Chu, D. E. Atkins, E. E. Swartzlander, and E. L. Ritman, *Comput. Biomed. Res.* **12**, 17 (1979).
15. H. H. Barrett and W. Swindell, *Proc. IEEE* **65**, 89 (1977).
16. E. W. Hansen and J. W. Goodman, *Proc. Soc. Photo-Opt. Instrum. Eng.* **231**, 222 (1980).
17. J. Holer, *Opt. Commun.* **29**, 22 (1979).
18. P. Edholm, L. G. Hellstrom, and B. J. Jacobson, *Phys. Med. Biol.* **23**, 90 (1978).
19. H. Platzer and H. Glunder, "Tomogram Reconstruction by Holographic Methods," in *Holography in Medicine and Biology*, G. V. Bally, Ed. (Springer, New York, 1979), pp. 117-123.
20. A. F. Gmitro, J. E. Greivenkamp, W. Swindell, H. H. Barrett, M. Y. Chiu, and S. K. Gordon, *Opt. Eng.* **19**, 260 (1980).
21. H. H. Barrett, *Opt. Lett.* **7**, 248 (1982).
22. I. M. Gelfand, M. I. Graev, and N. Ya. Vilkun, *Generalized Functions* (McGraw-Hill, New York, 1966), Vol. 5.
23. S. D. Shapiro, *Comput. Graph. Image Process.* **8**, 219 (1978).
24. J. Sklansky, *IEEE Trans. Comput.* **COM-27**, 923 (1978).
25. J. W. Goodman, "Linear Space Variant Optical Processing," in *Optical Data Processing, Fundamentals*, S. Lee, Ed. (Springer, New York, 1980).
26. R. J. Marks II, J. F. Walkup, M. O. Hagler, and T. E. Kride, *Appl. Opt.* **16**, 739 (1977).
27. D. Casasent and D. Psaltis, *Appl. Opt.* **15**, 1795 (1976).
28. D. Casasent and D. Psaltis, "Deformation Invariant, Space-Variant Optical Pattern Recognition," in *Progress in Optics*, Vol. 18, E. Wolf, Ed. (North Holland, Amsterdam, 1978).
29. G. Tydas, *Radiation and Propagation of Electromagnetic Waves* (Academic, New York, 1969), Chap. 8.
30. G. Eichmann, R. Stuhl, and R. Mammone, *Proc. Soc. Photo-Opt. Instrum. Eng.* **209**, 24 (1979).

Estimation of Two Closely Spaced Frequencies Buried in White Noise Using Linear Programming

Jaroslav Keybl and George Eichmann
Department of Electrical Engineering
The City College of the City University of New York
New York, N. Y. 10031

ABSTRACT

Linear programming is used to estimate the spectrum of two sinusoids signals closely-spaced in frequency buried in deep white gaussian noise by employing a-priori knowledge of the spectrum. The method will be illustrated by a number of examples.

INTRODUCTION

In the calculation of the spectrum of a discrete-time signal consisting of two sinusoids with closely-spaced frequencies embedded in white gaussian noise, problems arise when only a small portion of the signal is available. This is known as windowing of the data and becomes evident as "leakage" in the spectral domain, i.e. energy in the main lobe of a spectral response "leaks" into the sidelobes obscuring other present spectral responses. There are many methods used to estimate the spectrum. The periodogram [1] performance is poor for short data lengths. The Blackman-Tukey method [2] is also hampered by spectral distortion. The Burg method [3], a high frequency-resolution technique, even in the absence of noise, yields spectral line splitting. The recent algorithm of Cadzow [4] outperforms the Burg method in a low noise environment.

In this paper, we estimate the spectrum of two sinusoidal signals closely-spaced in frequency by employing a-priori knowledge of the Fourier spectrum of the signal in the form of linear inequalities. The advantage of imposing constraints in spectral restoration process has been pointed out [5]. In the linear programming formulation, there are many solutions. Here, we select a solution which minimizes the l_1 norm of the Discrete Fourier transform(DFT) of the measured signal and its estimate.

DETERMINATION OF THE SPECTRUM

We assume that the signal estimate $s(k)$ can be represented by a weighted sum of past signals

$$s(k) = -\sum_{i=1}^p a(i)s(k-i) + e(k) \quad k = 1, 2, \dots, m \quad (1)$$

where $a(i)$ are unknown weighting coefficients and $e(k)$ is an error term. This expression is a linear predictor[6]. By taking

WA18-2

the DFT of Eq.(1), we have

$$S(n) = - \sum_{i=1}^p a(i) \exp[-j2\pi i n/N] S(n) + E(n) \quad 1 \leq n \leq N \quad (2)$$

where $S(n)$ and $E(n)$ are the DFT of $s(k)$ and $e(k)$, respectively. Because $S(n)$ and $E(n)$ are complex, and using Euler's formula, we rewrite Eq.(2) as

$$S_R(n) = - \sum_{i=1}^p a(i) [S_R(n) \cos v + S_I(n) \sin v] + E_R(n) \quad (3a)$$

$$S_I(n) = - \sum_{i=1}^p a(i) [S_I(n) \cos v - S_R(n) \sin v] + E_I(n) \quad (3b)$$

where $v = 2\pi i n/N$. We have $2N$ equations with p unknowns. When the signal is real, we can reduce it to N equations with p unknowns. Assuming that two sinusoidal signals are present, it can be shown, the $a(i)$ coefficients are the pole coefficients of the z -transform of two sinusoidal signals. Therefore, $p = 4$ and the range of the $a(i)$'s are

$$-4 \leq a(1) \leq 4, \quad -2 \leq a(2) \leq 6, \quad a(3) = a(1), \quad a(4) = 1 \quad (4)$$

Eq.(4) adds 6 more equations, to the N equation generated from Eq.(3), for a total of $N + 6$ equations used in the linear programming formulation. From Eq.(3) and (4), we can now solve for the $a(i)$ coefficients by minimizing the l_1 norm of the error $-E(n)$.

NUMERICAL RESULTS

To test this method we have used the time series

$$s(k) = A_1 \cos(2\pi f_1 k) + A_2 \cos(2\pi f_2 k + \phi) + w(k) \quad (5)$$

with $1 \leq k \leq N$ and $w(n)$ as white Gaussian noise with zero mean and variance σ^2 . The two sinusoidal frequencies are normalized so that $f = 0.5$ corresponds to the Nyquist rate. The individual sinusoidal signal-to-noise ratio's (SNR) is given by $20 \log (A_k / \sqrt{2} \sigma)$ for $k = 1, 2$. For all of our examples, we chose the signal amplitudes $A_1 = \sqrt{2}$, $A_2 = \sqrt{20}$ and the signal frequencies $f_1 = 0.2168$ and $f_2 = 0.2245$. In two examples we introduced a forty-five degree phase difference between the two sinusoids.

In Fig. 1, the variance of of the noise $\sigma^2 = 0.94$ and $M = 192$. This corresponds to 0.54 db SNR on the weaker signal and an time-bandwidth product (TBP) of 1.50. The spectrum calculated using the periodogram shows random fluctuations and it resolves both frequencies. Our method yields peaks at frequencies $f_1 = 0.2168$ and $f_2 = 0.2265$ which shows the ability to resolve the frequencies with a small error in such a low SNR environment without fluctuation. It should be pointed out that the reason both peaks are of equal amplitude is because we are calculating the poles which make the function blow up. In Fig. 2, we introduce a forty-five degree phase shift between the

sinusoids. Here the SNR is - 0.26 db. We see that there is no line splitting as is often the case with other algorithms. The peaks occur at $f_1 = 0.2207$ and $f_2 = 0.2265$, again resolving the frequencies in a low SNR environment. Fig. 3 shows the estimated results when the SNR is - 14.2 db. The frequencies $f_1 = 0.2188$ and $f_2 = 0.2305$ yield good estimates for a very low SNR. In Fig. 4, we added a forty-five degree phase shift with a SNR - 15 db. We see that the periodogram is unable to resolve the two frequencies but our method peaked at $f_1 = 0.2188$ and $f_2 = 0.2324$. Again we note that there is no line splitting.² In the next set of experiments, the TBP is reduced. In Fig. 5, the SNR is 0 db and the TBP is 1.00. The periodogram shows peaks while our method gives the frequency estimates $f_1 = 0.2168$ and $f_2 = 0.2305$ showing that we get good quality estimates even when the number of samples is reduced. Finally, in Fig. 6, we plot the spectrum for a SNR of - 14.5 db. The estimated frequencies are $f_1 = 0.2168$ and $f_2 = 0.2363$. The larger frequency error can be attributed to the very low SNR and the small TBP environment.

SUMMARY

We have shown a new method for determining the spectrum of two sinusoidal signals closely-spaced in frequency, where the weaker signal has a very low SNR. We show that the method is not affected by spectral line splitting when the two sinusoids have a phase difference of 45 degrees. Computer generated results of this method have been presented.

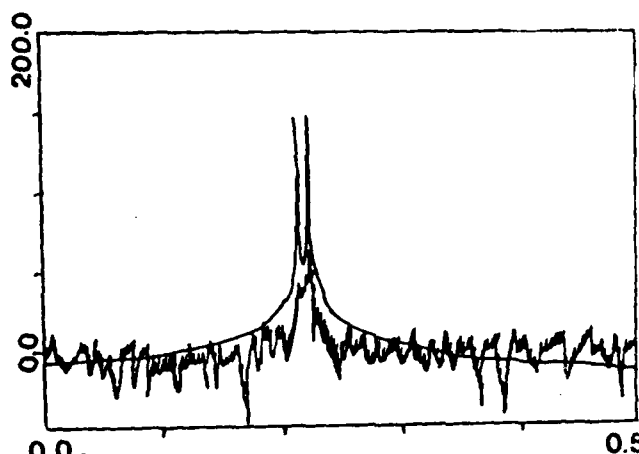
ACKNOWLEDGEMENT

This work is supported in part by a grant from the Air Force Office Scientific Research AFOSR 81-0169 and a contract from the Rome Air Development Center F19628-80-C-0095

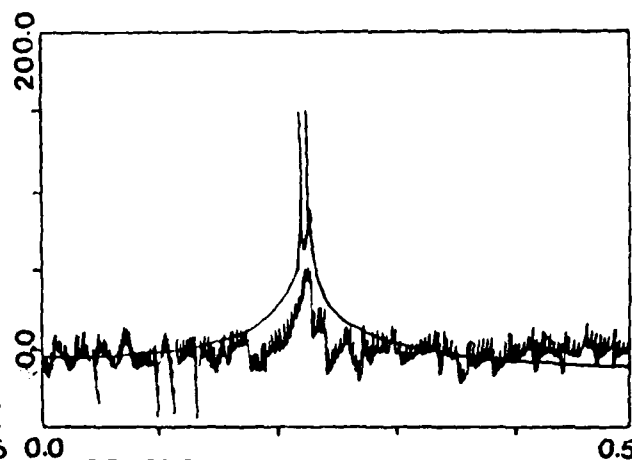
REFERENCES

- [1] S.M. Kay and S.L. Marple, Jr., "Spectrum Analysis - A Modern Perspective", Proc. IEEE, vol. 69, pp. 1320-1419, Nov. 1981.
- [2] R.B. Blackman and J.W. Tukey, The Measurement of Power Spectra From the Point of View of Communications Engineering (New York: Dover 1959)
- [3] A. Papoulis, "Maximum Entropy and Spectral Estimation: A Review," IEEE Trans. Acoust. Speech Signal Proc., vol. ASSP-29, pp. 1176-1186, 1981
- [4] J.A. Cadzow, "High Performance. Spectral Estimation - A New ARMA Method", IEEE Trans. Acoust. Speech Signal Proc., vol. ASSP-28, 524-529, Oct. 1980.
- [5] R.J. Mammone and G. Eichmann, "Restoration of the discrete Fourier spectra using linear programming", J. Opt. Soc. America, vol. 72, No. 8, August 1982
- [6] J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, vol. 63, pp. 561-580, April 1975.

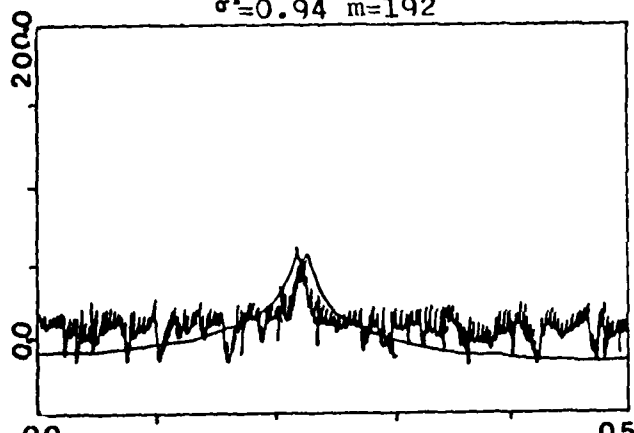
WA18-4



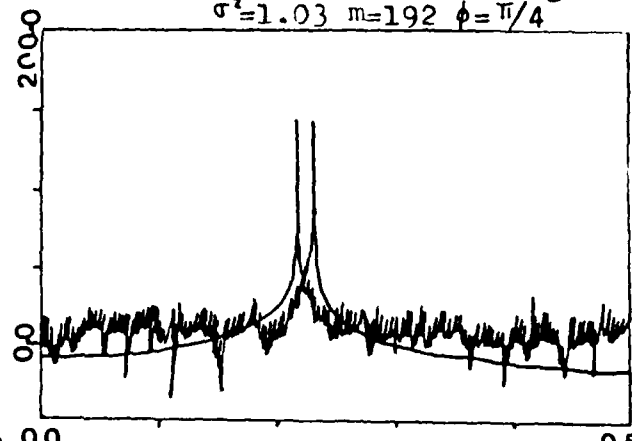
SPECTRUM-DB VS. FREQUENCY
Fig. 1. Spectrum using new
method and periodogram
 $\sigma^2=0.94$ $m=192$



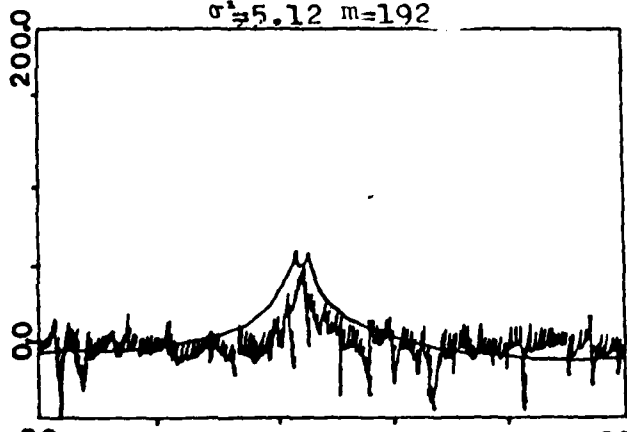
SPECTRUM-DB VS. FREQUENCY
Fig. 2. Spectrum using new
method and periodogram
 $\sigma^2=1.03$ $m=192$ $\phi=\pi/4$



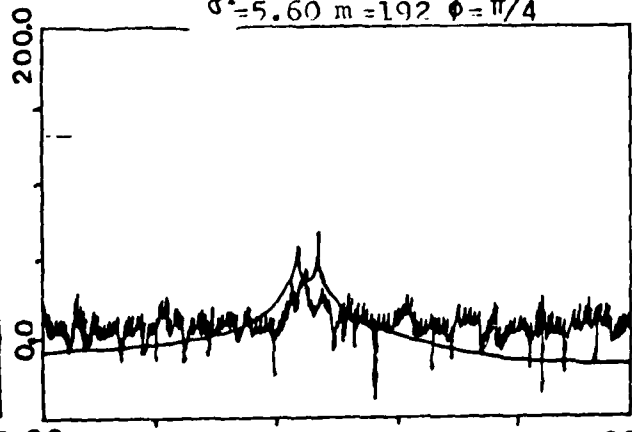
SPECTRUM-DB VS. FREQUENCY
Fig. 3. Spectrum using new
method and periodogram
 $\sigma^2=5.12$ $m=192$



SPECTRUM-DB VS. FREQUENCY
Fig. 4. Spectrum using new
method and periodogram
 $\sigma^2=5.60$ $m=192$ $\phi=\pi/4$



SPECTRUM-DB VS. FREQUENCY
Fig. 5. Spectrum using new
method and periodogram
 $\sigma^2=1.00$ $m=128$



SPECTRUM-DB VS. FREQUENCY
Fig. 6. Spectrum using new
method and periodogram
 $\sigma^2=5.27$ $m=128$ $\phi=\pi/4$

Restoration of discrete Fourier spectra using linear programming

R. Mammone* and G. Eichmann

Department of Electrical Engineering, The City College of the City University of New York, New York, New York 10031

Received December 31, 1981; revised manuscript March 15, 1982

A method of restoring the discrete Fourier transform (DFT) spectrum of a diffraction limited (DL) image from a narrow observation segment of the DL image is presented. The DL spectral restoration process is the dual of the more common DL image restoration process with the roles of the frequency and space reversed. Applications of a spectrum restoration include increasing the field of view of existing imaging systems and extracting precise frequency components of a large DL image by using only a small segment of the entire image. This method could also be employed for image data compression, which is of interest in digital video applications. Several differences between the implementations of the image and the spectrum restoration processes are described. The estimate is constrained to have an upper bound on the number of frequency components contained in the Fourier spectrum. The bound is the number of samples acquired at the Nyquist rate for the length of the image. The magnitude of the DFT spectrum is also bounded. These constraints define a large number of possible solutions. The desired solution is then selected such that the distance, defined in a function theoretic sense, between the measured and the estimated images is an optimum. A number of such measures are investigated. Numerical experiments show that this approach yields results that are highly immune to measurement noise.

INTRODUCTION

The finite aperture of any physical imaging system eliminates the high spatial-frequency components of the object from appearing in the image. The lack of high frequency detail results in a loss of resolution in the observed image. It has been shown that, for an object of finite extent, an exact restoration of the object from the diffraction-limited (DL) image is possible. The restoration of the DL object can be viewed as a continuation of the spatial Fourier spectrum beyond the spatial-cutoff frequency imposed by the DL system. Numerical methods of DL image restoration are highly unstable in the presence of measurement noise. The restoration process can be stabilized by imposing additional constraints.¹ Many restoration methods are available, such as those of Frieden,² Schell,³ Biraud,⁴ Jansson,⁵ and Burg.⁶ Recently, Howard⁷ demonstrated a computationally inexpensive method using a least-squares approach. This approach was further elaborated on by Rushforth *et al.*⁸ A new method of obtaining an increase of image resolution using linear programming techniques was recently given by Mammone and Eichmann.⁹

In this paper, the dual of the DL image-restoration problem is considered. Here the extrapolation of a finite segment of the DL (i.e., spatial bandlimited) image data in the presence of measurement noise is presented. The restoration of the discrete Fourier transform (DFT) spectrum may be interpreted as an extrapolation of the truncated spatial image. Applications of spectrum restoration include increasing the field of view of existing imaging systems and extracting exact frequency components of a large DL image when only a smaller image segment of the entire image is available. This reduction would be also of interest in image data compression applications, such as the transmission of digital video images. Although spectral restoration yields the same mathematical

formulation as that obtained for DL image restoration, with the role of space and spatial frequency reversed, there are significant differences as well. Here the unknowns are Fourier coefficients, which are complex numbers. This fact appears to double the numerical complexity of the problem. Also, the circulant convolutional operator that results from approximating the exact Fourier transform with a DFT spectrum leads to additional considerations. Finally, although the restored DL object is nonnegative, the restored DFT spectrum need not be real or nonnegative. Thus the powerful nonnegativity constraint used in the DL image restoration cannot be directly used.

The motivation of this research is to provide high resolution frequency estimates of data available only on an incomplete observation length by using the computationally efficient fast-Fourier-transform algorithms. To simplify the discussion, one-dimensional images are considered. For a DL (band limited) image truncated to lie within an interval S , the spatial frequency resolution is limited by the uncertainty principle.¹⁰ The lack of spectral resolution is related to the infinite limits of integration in the definition of the Fourier transform. Since the image is available over a finite observation length, the calculated Fourier transform is distorted. It is convolved with a sinc function with main lobe width D_x . For the case of two different sinusoidal images, the uncertainty principle states that the difference in the frequency between the images cannot be less than D_x ; otherwise there will be a significant overlap of the transforms. This overlap would not permit the two separate responses to be distinguishable. Thus the frequency resolution is limited to

$$D_x \geq 1/S, \quad (1)$$

In many practical situations, the image is not available for an interval of sufficient length S for the desired frequency res-

olution D_r . It is then desirable to process the spectrum of the observed signal such that the spectral resolution is greater than that given by formula (1). If there is no source of measurement error, it is well known¹ that the spectrum can be exactly restored. However, there is always noise and/or error present because of the finite numerical precision necessary to carry out the computation, if for no other reason. Therefore noise or measurement error must be considered as part of the formulation of the problem. In this case, the restoration process becomes highly unstable.

It has been found² that the imposition of *a priori* constraints, such as nonnegativity of the estimate, will stabilize the restoration process. Since the maximum number of independent equations that can be generated is equal to the number of spatial samples r required at the Nyquist rate, we shall restrict the spectrum-restoration method to situations in which the number of frequency components is less than or equal to r . This limit on the dimension of the estimate has been obtained in several other studies.¹⁰⁻¹² Further, the sinusoidal frequency components will be assumed to be harmonics of the same fundamental frequency. The latter assumption is not necessary but is made to facilitate the use of the fast-Fourier-transform (FFT) approximation of the Fourier spectrum. The periodic frequency spectrum is obtained by zero-padding the truncated space sequence such that the total number of elements is L , an integer that is a power of 2. This step allows a finer frequency scale to be generated.

The optimal data fitting employs linear programming (LP) techniques. The LP method provides both cost- and time-effective ways of selecting the optimal r -tuple estimate. LP methods have been used previously for the general image-restoration problem.^{9,13,14} The advantage of using the LP techniques for the unstable DL image restoration problem has also been addressed.⁹ In this paper we demonstrate the advantage of this approach for the rank-deficient spectrum-restoration problem. Since the rank, the number of independent equations, is less than the number of unknowns, many solutions exist. The deficiency of the rank occurs because of the DFT low-pass filter effect that eliminates all but r discrete frequency components. This effect is in contrast to the linear discrete-convolution case,⁹ which yields a system of equations that has a full rank. However, because the adjacent equations are almost identical, when a finite arithmetic machine representation is used, the system of equations is ill posed, i.e., nearly rank deficient.

FORMULATION

For the following discussion, it will be helpful to define negative spatial and frequency samples. The first $(L/2 + 1)$ elements of the sequence f_n or F_k correspond to the positive spatial or frequency samples, respectively, in ascending order, and the remaining $(L/2 - 1)$ elements correspond to the negative spatial or frequency samples, respectively, in descending order. This is consistent with the idea of extending the sequences in a periodic manner. The spatial truncation operation, the model for a short-observation segment, is characterized by multiplication of the spatial sequence by the unit rectangle sequence R_r , where r is the number of adjacent unity elements centered about the zero element and with all other elements are equal to zero. The DFT of the rectangle

sequence is real and even, since R_r is real and even, and is similar to a sampled-sinc function. Let the column vector h be defined as a vector whose elements are

$$h_i = \text{DFT}\{R_r\} \quad (2)$$

We shall use the circulant convolution theorem¹ to examine the relationship between the truncated and the exact sequences. The circulant convolution matrix for the discrete spatial truncation operator R is given by the circulant matrix

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 h_0 & \cdots & h_{L-1} \\ h_{L-1} & h_0 h_1 & \cdots & h_{L-2} \\ \vdots & \vdots & \ddots & \vdots \\ h_1 & h_2 h_1 & \cdots & h_0 \end{bmatrix}, \quad (3)$$

and therefore

$$G = \mathbf{H}F + N, \quad (4)$$

where G and F are the DFT's of the measured and the actual sequence vectors, respectively, and N is the complex error (noise) vector. It is well known¹⁵ that the eigenvalues of a circulant matrix \mathbf{H} are the coefficients of the DFT of the first row of the circulant matrix and the eigenvectors are the DFT basis vectors. Therefore the rank, which is equivalent to the number of nonzero eigenvalues, of \mathbf{H} is equal to the number of nonzero elements r in the rectangle sequence R_r . Since the vectors G , F , and N are complex, Eq. (4) represents $2L$ real scalar equations. But all the spatial sequences are real, and therefore the DFT sequences must have symmetry. The symmetry in turn introduces redundancies into Eq. (4). The elimination of this redundancy reduces the $2L$ to L real equations. In order to obtain these L equations we must examine some properties of the DFT.

Similar to the properties of the Fourier transform, the DFT of a real sequence possesses an even real and an odd imaginary part. Since we wish to use the computationally efficient radix two-FFT algorithm, the sequences must consist of an even number of points. The existing symmetry can be displayed in the following way. There is one element for the dc, the zeroth component and $(L/2 - 1)$ negative as well as positive elements, and one remaining, the $L/2$, element. This can be seen by noticing that the $L/2$ th row of the DFT matrix is a sequence of alternating positive and negative unities. Thus the center and the dc frequency components are always real. The remaining $(L - 2)$ elements consist of complex conjugate pairs between the first and the last elements and the second and the second-to-last elements, and so on. There are $(L/2 + 1)$ distinct real and $(L/2 - 1)$ imaginary elements. We form the concatenated sequence of the distinct elements. Let the real and the imaginary parts of the i th element be denoted by the subscripts Ri and Im , respectively;

then

$$\begin{aligned} \hat{G} &\approx [G_{R0}, G_{R1}, \dots, G_{R(L/2+1)}, G_{Im(L/2-1)}, \dots, G_{Im1}]^T, \\ \hat{F} &\approx [F_{R0}, F_{R1}, \dots, F_{R(L/2+1)}, F_{Im(L/2-1)}, \dots, F_{Im1}]^T, \\ \hat{N} &\approx [N_{R0}, N_{R1}, \dots, N_{R(L/2+1)}, N_{Im(L/2-1)}, \dots, N_{Im1}]^T, \end{aligned} \quad (5)$$

where the superscript T stands for the transpose operation. The relation Eq. (4) can now be written in reduced form

$$\hat{G} = \hat{\mathbf{H}}\hat{F} + \hat{N}, \quad (6)$$

where, if we decompose the degradation matrix, then

$$\mathbf{H} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}, \quad (7)$$

where \mathbf{A} is a $(L/2 + 1) \times (L/2 + 1)$, \mathbf{B} is a $(L/2 + 1) \times (L/2 - 1)$, \mathbf{C} and \mathbf{D} are $(L/2 - 1) \times (L/2 - 1)$ submatrices of \mathbf{H} and \mathbf{e} , and \mathbf{p} are unused column vectors. If we denote the null vector and matrix by 0 and $\mathbf{0}$, respectively, and define

$$\mathbf{B} = \{\mathbf{0}\mathbf{B}^R\mathbf{0}\}, \quad (8)$$

$$\hat{\mathbf{C}} = \{\mathbf{0}\mathbf{0}\mathbf{C}\}, \quad (9)$$

$$\mathbf{D} = \{\mathbf{0}\mathbf{0}\mathbf{D}^R\}, \quad (10)$$

where the superscript R denotes reverse order, then

$$\hat{\mathbf{H}} = \begin{bmatrix} \mathbf{A} + \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{D}} - \hat{\mathbf{C}} \end{bmatrix}. \quad (11)$$

Thus Eq. (6) represents L real equations in the $2L$ variables $\hat{\mathbf{F}}$ and $\hat{\mathbf{N}}$. Here $\hat{\mathbf{H}}$ denotes the space truncation operator in the DFT domain. This underdetermined system of equations has many solutions. For example, the pseudo-inverse solution might be used,¹⁶ although this would produce an unconstrained estimate. In order that we may be able to impose constraints on and simultaneously choose an optimal solution, that is, one that is close in some sense to the measured image, we use a method of optimization known as linear programming.

CONSTRAINED OPTIMAL SOLUTIONS

The basic linear-programming (LP) problem is formulated in the following way¹⁷:

$$\begin{aligned} &\text{Minimize the cost function} && c^T x \\ &\text{Subject to} && \mathbf{A}x \leq b, \\ &&& x \geq 0, \end{aligned} \quad (12)$$

where c and x are the known cost and unknown M -dimensional variable vectors, respectively, b is an r -dimensional constraint vector, and \mathbf{A} is an $r \times M$ constraint matrix. The most often used LP technique is the simplex method.¹⁷ The output obtained from a simplex LP algorithm will correspond to one of the following three situations: There is no feasible solution; that is, all the constraints cannot simultaneously be satisfied; or the optimal solution is feasible but is unbounded; the third possibility is that the solution is feasible and bounded, and then an optimal solution is provided.

The first step in the simplex method is to convert the inequality constraints into equalities. This step is accomplished by introducing additional variables, called slack variables. The resulting system of equations will have more unknowns than equations, leading to an underdetermined system of equations. From the many possible solutions the simplex method seeks a solution that minimizes the cost function. The basic premise of LP is that the solution vector x will consist of at most r nonzero elements, where r equals the number of independent equations, and the remaining $M - r$

elements will be equal to zero. In a vector-space interpretation of LP, the inequalities define intersecting hyperplanes, which form a region of possible solutions. This region is an r -dimensional polygon or simplex. The simplex will always lie in the positive octant of the vector space. The cost function also defines a hyperplane for a fixed-cost value. For this value to be minimum, the cost-function hyperplane must lie on an edge of the simplex. The coordinates of each edge contain at most r numbers. The simplex method is an iterative technique, which begins with a feasible solution at an arbitrary edge of the simplex and progresses to an adjacent edge in such a way as to yield a reduction in the cost value until no further reduction is possible. The coordinates of the resulting edge yields the optimal solution.

In a LP formulation of the spectral-estimation problem, we now take Eq. (6) to be part of the constraint matrix. To guarantee that all the variables are nonnegative, a necessary requirement in LP, we next define the vectors in Eq. (6) in terms of the nonnegative component vectors $\hat{\mathbf{F}}^+$, $\hat{\mathbf{F}}^-$, $\hat{\mathbf{N}}^+$, and $\hat{\mathbf{N}}^-$ such that

$$\hat{\mathbf{F}} = \hat{\mathbf{F}}^+ - \hat{\mathbf{F}}^- \quad (13)$$

and

$$\hat{\mathbf{N}} = \hat{\mathbf{N}}^+ - \hat{\mathbf{N}}^-. \quad (14)$$

If we examine the sum

$$\hat{\mathbf{N}}^+ + \hat{\mathbf{N}}^- = \|\hat{\mathbf{G}} - \hat{\mathbf{H}}\hat{\mathbf{F}}\|, \quad (15)$$

we observe that Eq. (15) is the l_1 norm of the error between the spectrum of the measured signal $\hat{\mathbf{G}}$ and the spectrum of the spatially truncated image estimate $\hat{\mathbf{F}}$. This is the error measure to be minimized. Thus the LP formulation of the l_1 spectral estimation problem is

$$\begin{aligned} &\text{Minimize} && \hat{\mathbf{N}}^+ + \hat{\mathbf{N}}^- \\ &\text{Subject to} && \hat{\mathbf{G}} = \hat{\mathbf{H}}(\hat{\mathbf{F}}^+ - \hat{\mathbf{F}}^-) + \hat{\mathbf{N}}^+ - \hat{\mathbf{N}}^-, \\ &&& \hat{\mathbf{F}}^+ \leq \hat{\mathbf{F}}_{\max}^+, \\ &&& \hat{\mathbf{F}}^- \leq \hat{\mathbf{F}}_{\max}^-, \quad \hat{\mathbf{F}}^+, \hat{\mathbf{F}}^-, \hat{\mathbf{N}}^+, \hat{\mathbf{N}}^- \geq 0. \end{aligned} \quad (16)$$

The l_1 norm is not the only measure of error that can be minimized. The l_{∞} norm, or maximal deviation, can also be employed. This can be seen by defining a scalar E to be the absolute value of the maximum element of the error vector given by Eq. (15). Thus the LP formulation of the l_{∞} spectral-estimation problem is

$$\begin{aligned} &\text{Minimize} && E \\ &\text{Subject to} && \hat{\mathbf{G}} \leq \hat{\mathbf{H}}(\hat{\mathbf{F}}^+ - \hat{\mathbf{F}}^-) + E\mathbf{e}_n, \\ &&& \hat{\mathbf{G}} \geq \hat{\mathbf{H}}(\hat{\mathbf{F}}^+ - \hat{\mathbf{F}}^-) - E\mathbf{e}_n, \\ &&& \hat{\mathbf{F}}^+ \leq \hat{\mathbf{F}}_{\max}^+, \\ &&& \hat{\mathbf{F}}^- \leq \hat{\mathbf{F}}_{\max}^-, \quad \hat{\mathbf{F}}^+, \hat{\mathbf{F}}^-, E \geq 0, \end{aligned} \quad (17)$$

where \mathbf{e}_n is an L -dimensional unit vector.

The l_2 norm, or the sum of the squares of the error, can also be minimized by using a formulation similar to that of the l_1 norm. Here quadratic programming methods are used. Quadratic programming techniques provide an estimate with at most r positive elements, with all other elements as zero. In general, quadratic programming techniques use computationally efficient LP methods but on larger augmented matrices. These methods are well known, and many pub-

lished computer routines are available.^{18,19} There is usually more than one least-squares estimate to an undetermined system of equations such as Eq. (4). For example, the pseudo inverse solution¹⁵ provides an unconstrained least-squares estimate with the lowest l_2 norm as well as the minimum-squares error. Howard⁷ has demonstrated a constrained least-squares approach for the dual problem incorporating a penalty function to force the solution to become nonnegative. Rushforth *et al.*⁸ developed this approach for an ill-posed DL image restoration problem.

The estimates obtained by optimizing the l_1 , l_2 , and l_{∞} norms are maximum-likelihood estimates when the noise is modeled as a random variable with uniform, Gaussian, or exponential probability density functions (pdf's), respectively.²⁰ It was found, however, that for the low noise situation of interest, the pdf of the noise did not appreciably affect the estimates. In terms of increased frequency resolution, the l_1 norm consistently offered the best estimate, irrespective of the noise pdf. This fact can be attributed to the relatively fewer numerical operations needed as well as to the strongest bounds on the individual spectral components for the l_1 estimate. The constrained minimal l_2 norm using quadratic programming techniques requires many more numerical operations. Also, optimizing the l_2 norm is equivalent to minimizing the total noise spectral energy. Whereas this is a desirable feature in general, for the purpose of increased frequency resolution it tends to be counterproductive. The minimization of the global spectral-error noise-energy distribution tends overly to smooth the estimate, leading to a decrease in frequency resolution.

NUMERICAL RESULTS

The observed spectrum is a smoothed version of the desired spectrum. Figure 1 illustrates the inadequacy of simply taking the FFT of a truncated sequence. Here, as well as in all subsequent figures, the magnitude of the Fourier spectrum is plotted. The dotted line represents a 32-point FFT of 23 samples of a $\cos(u_L x)$ sequence, where u_L is the fundamental frequency $2\pi/32$. Here only seven data samples have been replaced by zeros. The solid line illustrates the exact spectrum of the full 32-sample sequence. The lack of resolution conceals the fact that only the fundamental frequency is present. The size of the FFT as well as the number of samples acquired is known *a priori*. This information, together with the bounds on the amplitude spectrum, is used to estimate the

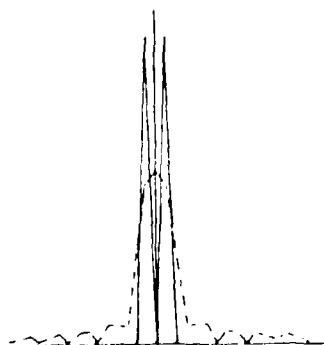


Fig. 1. The 32-point FFT of 23 sample points of $\cos(u_L x)$ (dotted-dashed curve) and the exact DFT spectrum (solid line).

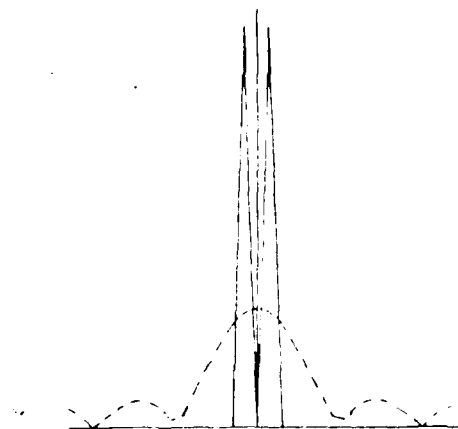


Fig. 2. The restoration of the FFT of $\cos(u_L x)$ from the 32-point FFT of 7 spatial domain sample points (dotted-dashed curve) and the estimated spectrum (solid line).

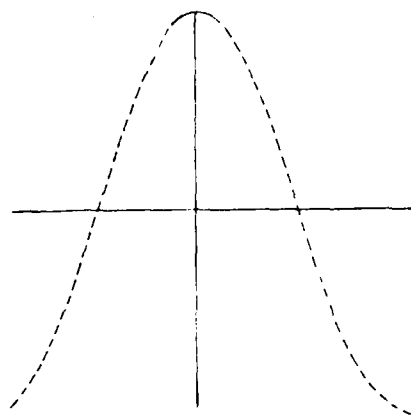


Fig. 3. Extrapolation of $\cos(u_L x)$ in the spatial domain (dashed line curve) from 7 sample points (solid line).

actual spectrum. Figure 2 illustrates the performance of the l_1 norm when only 7 of the 32 samples of the fundamental cosinusoid are available. The dotted line curve represents the overly smooth estimate of the direct FFT spectrum. The minimal l_1 estimate, the dotted line, is close to the original spectrum. The dotted line curve lies completely on the solid-line curve. Thus a complete restoration is possible even when only 7 of 32 samples are available. Figure 3 illustrates the extrapolated space domain sequence corresponding to Fig. 2. The acquired samples are illustrated by the solid line segment and the extrapolated curve by the dashed-line curve.

The method stands up quite robustly to measurement noise. Figure 4 illustrates the restoration of the fundamental cosinusoidal waveform from 15 samples of a 32-point FFT with white Gaussian noise added. The variance of the noise is one tenth of the amplitude of the function. Thus an increase of the order of a factor of 2 in resolution is obtained even in the presence of significant measurement noise. The corresponding spatial extrapolation is depicted in Fig. 5. Here the incomplete observation is shown by the dotted line curve. The extrapolated waveform (the chain-dotted curve) follows the exact curve (solid line) closely. In many applications it

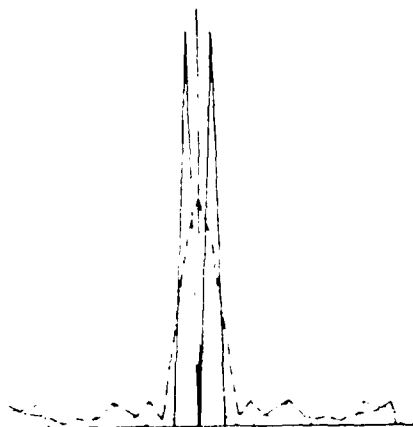


Fig. 4. Restoration of $\cos(\pi x)$ from 15 spatial sample points (dotted-dashed curve) with white Gaussian noise. The noise variance is 0.1. The restored spectrum is depicted by a solid line.

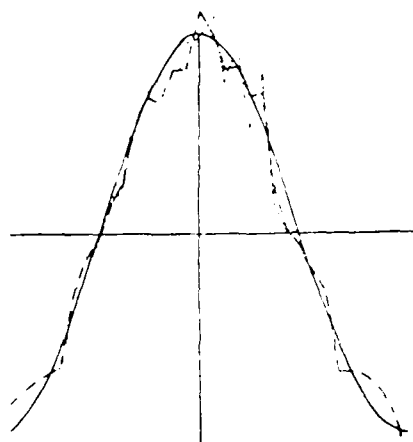


Fig. 5. Extrapolation of $\cos(\pi x)$ (chain-dotted curve) in the spatial domain corresponding to frequency domain plot of Fig. 4. The dotted-dashed curve indicates the incomplete observation with measurement error. The extrapolated curve (solid line) follows the actual waveform (dotted line) closely.

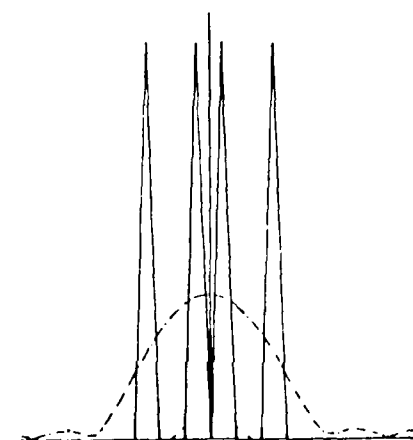


Fig. 6. Restoration of $\cos(\pi x) + \cos(5\pi x)$ using a 32-point FFT from 11 spatial domain sample points. The DFT is depicted by a dotted-dashed curve, and the restored waveform is shown by a solid curve.

is desired to resolve two frequency components spaced close together. Figure 6 demonstrates the performance of the l_1 norm in this case. Here the first and the fifth harmonics are sampled with 11 of the possible 32 samples. The direct application of the FFT provides the blurred spectrum indicated by the chain-dotted curve. The estimated Fourier spectrum is represented by the solid line; again, the estimate follows the exact spectrum so closely that the curves are indistinguishable.

SUMMARY AND CONCLUSIONS

A new method of extrapolating a DL image from a measured spatially truncated image is presented. The method directly addresses the issue of the finite degree of freedom of the DL image. The feasible image estimates are constrained to have an upper bound on the number of frequency components comprising the image. A geometric approach has been taken, that is, various error measures were optimized subject to given constraints. Three error norms were discussed—the l_1 , l_2 , and l_∞ norms. This method permits the addition of other linear constraints on the desired image. That is, any inequality possessing a linear combination of the unknown samples on one side of the inequality and a scalar value on the other side can also be added. The inclusion of constraints of this type has been found to stabilize the otherwise unstable problem. Numerical results demonstrate the increase in resolution as well as its strength against measurement noise. In terms of resolution, the l_1 norm was found to give consistently the best spectral estimates.

ACKNOWLEDGMENTS

This paper is based on a portion of a dissertation submitted by R. J. Mammone to the City University of New York, in partial fulfillment of the requirements for the Ph.D. degree in engineering. This work was supported in part under grant AFOSR 81-0169 from the U.S. Air Force Office of Scientific Research and contract F19628-80-C-0095 from the Rome Air Development Center.

* Present address, Department of Electrical Engineering, Rutgers University, Piscataway, New Jersey 08854.

REFERENCES

1. D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions. Fourier analysis and uncertainty. I," *Bell Syst. Tech. J.*, **40**, 43-63 (1961).
2. B. R. Frieden, "Image enhancement and restoration," in *Picture Processing and Digital Filtering*, T. S. Huang, ed. (Springer-Verlag, New York, 1970).
3. A. C. Schell, "Enhancing the angular resolution of incoherent sources," *Radio Electron. Eng.*, **29**, 21-26 (1965).
4. Y. Brund, "A new approach for increasing the resolving power by data processing," *Astron. Astrophys.*, **1**, 124-127 (1969).
5. P. A. Jansson, R. H. Hunt, and E. K. Peyer, "Resolution enhancement of spectra," *J. Opt. Soc. Am.*, **60**, 596-599 (1970).
6. J. P. Burg, "Maximum entropy spectral analysis," presented at the Thirty-Seventh Annual Meeting of Society of Exploration Geophysicists, Oklahoma City, Oklahoma, 1967.
7. S. J. Howard, "Continuation of discrete Fourier spectra using a minimum negativity constraint," *J. Opt. Soc. Am.*, **71**, 819-824 (1981).

8. C. K. Rushtorth, A. E. Crawford, and Y. Zhou, "Least squares reconstruction of objects with missing high frequency components," *J. Opt. Soc. Am.* **72**, 204-211 (1982).
9. R. Mammone and G. Eichmann, "Superresolving image restoration using linear programming," *Appl. Opt.* **21**, 496-501 (1982).
10. A. Papoulis, *Signal Analysis* (McGraw-Hill, New York, 1977).
11. H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty. III: the dimension of the space of essentially time- and band-limited signals," *Bell Syst. Tech. J.* **41**, 1295-1336 (1962).
12. G. Toraldo de Francia, "Degrees of freedom of an image," *J. Opt. Soc. Am.* **59**, 799-804 (1969).
13. D. P. MacAdam, "Digital image restoration by constrained convolution," *J. Opt. Soc. Am.* **60**, 1617-1627 (1970).
14. N. D. A. Mascarinas and W. K. Pratt, "Digital restoration under a regression model," *IEEE Trans. Circuits Syst.* **CAS-22**, 252-266 (1975).
15. B. R. Hunt, "A matrix theory proof of the discrete convolution theorem," *IEEE Trans. Audio Electroacoust.* **AF-19**, 285-288 (1971).
16. A. Albert, *Regression and the Moore-Penrose Pseudoinverse* (Academic, New York, 1972).
17. G. Hadley, *Linear Programming* (Addison-Wesley, Reading, Mass., 1962).
18. H. P. Kunzi, H. G. Tzschach, and C. A. Zehnder, *Variational Methods of Mathematical Optimization* (Academic, New York, 1971).
19. C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Prentice-Hall, Englewood Cliffs, N. J., 1974).
20. S. M. Pizer, *Numerical Computational Mathematical Analysis* (Science Research Association, Chicago, Ill., 1975).

Two-dimensional optical filtering of 1-D signals

G. Eichmann and B. Z. Dong

One-dimensional signals have complementary attributes, their space and spatial frequency properties. Recently, another signal representation was introduced in optics, the Wigner distribution (WD) function, which allows the simultaneous display of the two attributes of a 1-D signal. In this paper optical implementation of a generalized space-spatial frequency (GSF) function is discussed. Special cases of the GSF representations are the WD, the radar ambiguity function (AF), the energy distribution function, the various pseudo-WD functions, the spectrogram, the local frequency and the local Doppler frequency spectrum. The GSF representation allows the simultaneous filtering of both the spatial and the spatial frequency content of a 1-D signal. The 2-D filtering of a 1-D signal allows sidelobe reduction of the WD and AF of quasi-periodic 1-D signals; display of the instantaneous spatial frequency content of a 1-D signal; space-spatial frequency extension of a desired portion of a 1-D signal as well as the usual matched filter detection. Experimental results using coherent space-integrating techniques for the realization of different GSF are presented.

I. Introduction

In the area of optics dealing with the optical processing of 1-D signals such as speech, radar, sonar, or general communication signals, the 2-D nature of the optical system has been viewed primarily as the means for providing the potential for greatly increased information throughput.¹⁻³ Thomas¹ has shown, for example, how a coherent optical system with a large space-bandwidth product in one dimension can be used to perform spectrum analysis of a 1-D signal with a time-bandwidth product of the order of a million. More recently, a large number of space-variant coherent optical processing operations, performed on a 1-D signal, have been added to the repertoire, such as frequency-variant spectral filtering,⁴ variable spatial magnifications, and geometrical distortions,^{5,6} as well as various 1-D signal transforms such as Mellin,^{7,8} Abel,⁸ and Laplace. An excellent tutorial on space-variant coherent optical processing is given by Walkup.⁹

Coherent optical processors can also be used to evaluate attributes of two identical or different 1-D signals such as convolution or correlation between the two signals. A particular signal representation, the

so-called radar ambiguity function (AF),¹⁰⁻¹² uses the whole output plane of a coherent Fourier processing system. Here one axis represents a time delay, the range information, while the other axis represents the Doppler frequency shift, a measure of the target motion. Another interpretation of the information displayed on the two axes is that one axis represents the correlation while the other axis represents the energy or power spectrum of the 1-D signal. Recently, another signal representation was introduced in optics, the Wigner distribution function (WD).¹³⁻¹⁵ The WD of a single 1-D signal represents the spatial variable on one axis and the energy or power spectrum on the other axis. Another interpretation of the WD, which is similar to the AF, is that the displayed information represents convolution on one axis and the energy or power spectrum on the other axis. The two signal representations are not independent. In fact, both the AF and the WD are special cases of a general class of representations of a 1-D signal.¹⁶ In both cases the 1-D signal is represented in a 2-D space. The two independent coordinates represent important physical information about the 1-D signal.

The two spatial axes, coherent Fourier optical implementations of the AF display and their modification, lead easily to 2-D spatial filtering of 1-D signals. The purpose of this paper is to explore various 2-D optical filtering schemes for 1-D signals. Such signals have applications in system modeling, bandwidth compression of speech or video signals as well as in the general understanding of the fundamental limitation for time-frequency representations.

The authors are with City College of the City University of New York, Department of Electrical Engineering, New York, New York 10031.

Received 15 March 1982.

0003-6935/82/173152-05\$01.00/0

© 1982 Optical Society of America.

II. Formulation

The starting point for 2-D filtering of 1-D signals is the generalized space frequency (GSF) representation of 1-D signals introduced by Cohen.¹⁶ This signal class is defined by the triple integral

$$C(x, u, H) = \frac{1}{2\pi} \int \int \int \exp[ju(x - zu - v)] \times H(v, z) g(x, z) dv dz, \quad (1)$$

where

$$g(x, z) = f(x + z/2) f^*(x - z/2), \quad (2)$$

where $f(x)$ is the 1-D signal, the $*$ represents complex conjugate, and $H(v, z)$ is an arbitrary kernel function. Particular kernel functions lead to different space-frequency representations of $f(x)$. Two of the most important GSF representations of the 1-D signal are the AF and the WD.

The AF is defined as

$$A(v, z) = \int_{-\infty}^{\infty} \exp(-jvz) g(x, z) dx \quad (3)$$

An alternate representation of the AF can be expressed in terms of the Fourier transform of $f(x)$:

$$F(u) = \int_{-\infty}^{\infty} \exp(-jux) f(x) dx \quad (4)$$

The AF can now be expressed in terms of the Fourier spectra as

$$A(v, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(jvz) G(u, u) du, \quad (5)$$

where

$$G(v, u) = F(u + v/2) F^*(u - v/2). \quad (6)$$

The WD, another GSF representation, is defined as

$$W(x, u) = \int_{-\infty}^{\infty} \exp(-juz) g(x, z) dz \quad (7)$$

Similar to the AF, the WD can be found from the Fourier spectra as

$$W(x, u) = \int_{-\infty}^{\infty} \exp(jvx) G(v, u) dv \quad (8)$$

The two representations are not independent. In fact the AF and the WD are related by a scaled 2-D Fourier transform:

$$A(v, z) = \frac{1}{2\pi} \iint \exp[-j(vx - uz)] W(x, u) dx du \quad (9)$$

The WD of a 1-D signal can be interpreted as the distribution of signal energy over space and spatial frequency. Thus a 2-D display of the WD represents the energy content of the waveform localized in space and spatial frequency. However, this interpretation is not always true since there are cases where the WD is not positive. One of the most important properties of the WD is that a shift in either space or spatial frequency of the 1-D signal leads to a proportional shift in the corresponding WD. This is not true for the AF where such shifts lead to an additional phase factor.¹⁷ In fact,

the magnitude of the AF is completely insensitive to shifts in space or spatial frequency. Detailed properties of both the AF and WD can be found in the literature.^{17,18}

Both the AF and the WD are special cases of the GSF distribution C . In fact, the GSF function C can be expressed either in terms of the AF as

$$C(x, u, H) = \frac{1}{2\pi} \iint \exp[ju(x - uz)] H(v, z) A(v, z) dv dz, \quad (10)$$

or in terms of the WD as

$$C(x, u, H) = \frac{1}{2\pi} \iint h(v - z, u - v) W(v, z) dv dz, \quad (11)$$

where

$$h(v, u) = \frac{1}{2\pi} \iint \exp[ju(x - vz)] H(v, z) dv dz \quad (12)$$

is a scaled 2-D Fourier transform of the kernel function H . The kernel function $h(x, u)$ represents a 2-D impulse response filter acting on the space-frequency dependent WD. Since the WD and the AF are related by 2-D Fourier transformation, the kernel function H acts as a 2-D filter on the AF.

An all-pass filtering function

$$H(v, z) = 1 \quad (13)$$

transforms the GSF function C into the WD. An impulsive kernel function

$$H(v, z) = 2\pi \delta(v - z) \delta(u - v) \quad (14)$$

transforms the GSF function C into the AF. Other filter functions have appeared in the literature. For example, the all-pass filter

$$H(v, z) = \exp[jv^2/(2z)] \quad (15)$$

transforms the GSF distribution C into an energy distribution function. This filter function was first proposed by Ribacek.¹⁹ The energy distribution function has properties similar to the WD. Another filter function is a Gaussian kernel

$$H(v, z) = \exp[-(v^2/a^2 + z^2/b^2)] \quad (16)$$

This function has been suggested by deBruijn²⁰ as an aid in the interpretation of the WD, a non-negative energy distribution. It can be shown that, by choosing the parameters a and b in a particular way, the modified WD will always be real and non-negative. Further, this type of GSF distribution is then consistent with Heisenberg's uncertainty principle.

III. Two-Dimensional Filtering of 1-D Signals

The GSF distribution can be interpreted therefore either as a generalized WD (GWD) or as a generalized AF (GAF). While the present discussion will deal with the GWD, without any loss of generality it can be carried over to a discussion of the GAF. Particular filtering of the 1-D signal $f(x)$ leads to a new signal $k(x)$. The WD of the new signal is related to the WD of $f(x)$ through a linear 2-D filter. This GSF filter is the WD of the 1-D filter function.

For example, the 1-D linear space-invariant filtering of $f(x)$ is represented as

$$k(x) = f(x) * l(x) = \int_{-\infty}^{\infty} f(y)l(x-y)dy, \quad (17)$$

where $l(x)$ is the spatial impulse response. The WD of the filtered signal $k(x)$ is

$$K = W_k(x,u) = \int_{-\infty}^{\infty} W_f(y,u)W_l(x-y,u)dy, \quad (18)$$

where W_l is the WD of the filter function. Comparing Eq. (11) with the present equation we note that the new WD is the GSF. The new WD may also be considered as a generalization of the WD of $f(x)$. Filtering the signal in the space domain also leads to filtering of the WD in the space domain.

Multiplication of the signal $f(x)$ in the space domain

$$k(x) = f(x)m(x) \quad (19)$$

leads to GWD:

$$W_k(x,u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W_f(x,u)W_m(x-u,y)dy \quad (20)$$

This equation represents 1-D filtering of the WD in the spatial frequency domain. Here again W_m is the WD of the multiplication function $m(x)$. This is another GWD of the signal $f(x)$.

Another type of filtering can be obtained by weighing the signal with a sliding window function

$$k(x) = f(x)m(x-y) \quad (21)$$

The GWD of the signal $k(x)$ is

$$W_k(x,u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W_f(x,t)W_m(x-y,u-t)dt \quad (22)$$

If we consider values of the GWD only on the $x=y$ axis, we obtain a new version GWD called a pseudo-WD. There are other filter operations that can be performed on either $f(x)$ or on $F(u)$ that will lead to GWD. By defining the filtering operations on the 2-D Fourier transformed WD, a GAF can also be found. Depending on the type of information needed, space-spatial frequency or space-spatial frequency delays, either GWD or GAF can be generated.

Another 2-D display of a 1-D signal is the short-space Fourier transform (SFT).²⁴ This transform is given as

$$Q(x,u) = \int_{-\infty}^{\infty} \exp(-juy)k(x,y)dy, \quad (23)$$

where $k(x,y)$ is the expression given by Eq. (21). The spectrogram $S(x,u)$ is obtained by taking the square of the magnitude of the SFT for all possible window positions:

$$S(x,u) = |Q(x,u)|^2 \quad (24)$$

However, the spectrogram can also be expressed as

$$S(x,u) = \frac{1}{2\pi} \iint_{-\infty}^{\infty} W_f(y,z)W_m(y-x,z-u)dydz \quad (25)$$

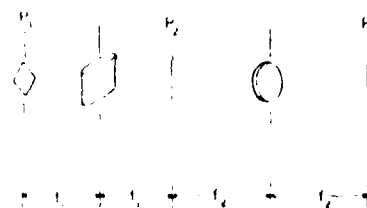


Fig. 1. Basic coherent optical arrangement to generate the AF and the modified WD. In plane P_1 , two 1-D signal masks are placed next to each other with the masks turned 90° apart. The mask assembly is rotated 45° relative to the lens symmetry axis. In plane P_2 the magnitude of the modified WD is generated, while in plane P_3 the magnitude of the AF is generated.

Therefore, the spectrogram is also a member of the GSF. In this case, a fully 2-D filter smooths the WD of the signal $f(x)$.

IV. Optical Realizations of 2-D Filtering of 1-D Signals

A starting point for the optical realizations of the 2-D filtering of 1-D signals is the large number of optical realizations for the AF. There are coherent space-integrating,^{10,11} coherent time-integrating,²² incoherent time-integrating,²³ coherent hybrid (time and space integrating),²⁴ coherent joint Fourier transform holography,²⁵ as well as other methods²⁶⁻²⁹ available to optically display the AF. While all the above implementations can be modified to be used as a starting point for the 2-D filtering of 1-D signals, here we discuss some experimental space-integrating coherent optical implementations of these filtering operations.

The basic coherent optical realization of the WD is shown in Fig. 1. The field in plane P_1 is

$$g(x,y) = f\{t/\sqrt{2}(x+y)\}f^*[1/\sqrt{2}(x-y)]. \quad (26)$$

In plane P_2 , we have

$$G(x,y) = \iint_{-\infty}^{\infty} g(x',y') \exp[jk/2f(xx')] \times \exp[-jk/2f(yy'-y'')]dx'dy', \quad (27)$$

where k is the wave number and f is the focal length of the cylindrical lens. Due to the cylindrical lens, we have a Fourier transform in the x direction and a Fresnel kernel transform in the other direction. The above integral can be expressed as

$$G(x,y) = \iint_{-\infty}^{\infty} W(x',y')\delta(x-x') \exp[-jk/2f(y-y')]dx'dy' \quad (28)$$

In plane P_3 a filtered version of the WD is generated. Using a spherical lens, the convolution operation can be mapped into a product of Fourier transforms. Thus in plane P_3 we have the AF multiplied by a quadratic phase factor. The magnitude of the product is the magnitude of the AF. The phase factor can be elimi-



a



b

Fig. 2. Outputs, in planes P_1 and P_2 of Fig. 1, of a 1-D mask containing five different spatial frequencies are given: (a) amplitude of the modified WD, (b) amplitude of the AF.

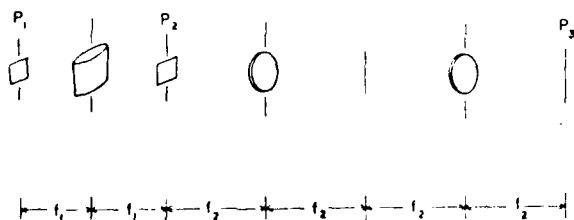


Fig. 3. Coherent optical implementation for the generation of the energy or Rihaczek distribution. The 1-D spatial Fourier spectrum illuminates the 1-D signal mask turned 90°. The two spherical lenses image the energy distribution.

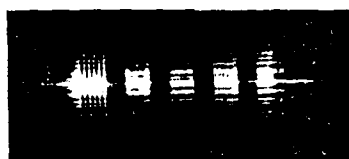
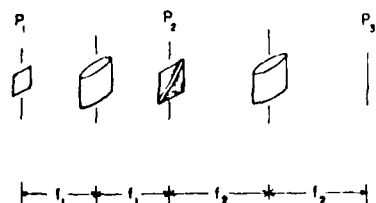
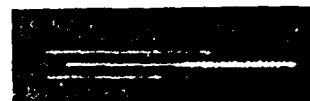


Fig. 4. Energy distribution of a 1-D mask containing five different spatial frequencies.

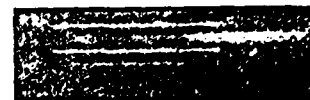


(Left) Fig. 6. Coherent optical implementation of the SFT. Here the running window function weighs the complex 1-D Fourier spectra of the signal. (Middle) Fig. 7. (a) Spectrogram of the double-sided frequency ramp. Because of the impulsive nature of the filtering this spectrogram is termed local spectra. (b) Local Doppler spectra of the double-sided frequency ramp. (Right) Fig. 8. (a) Local spectra. (b) Local Doppler spectra of a 1-D mask containing five different spatial frequencies.

nated using appropriate lenses.⁶ Figure 2 shows the magnitude of the filtered WD, filtered by a Fresnel kernel, and the magnitude of the AF of five different spatial frequencies. Both the filtered WD and its AF show a large number of sidelobes characteristic of grating patterns. Figure 3 shows the optical implementation of a third representation of the 1-D signal, the Rihaczek or energy distribution. Here the 1-D Fourier transform of the signal is used to illuminate the original 1-D signal mask turned 90°. The first



a



b



c



d

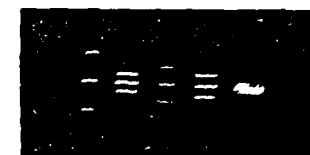
Fig. 5. Results of vertical-slit filtering of the input masks in plane P_1 of Fig. 1: (a) and (b) results of this filtering on a 1-D mask containing two spatial frequencies; (c) and (d) results of this filtering on a 1-D mask containing five spatial frequencies with and without de spectral components.



a



b



a



b

input is a phase mask. The transilluminated mask is in turn imaged using two spherical lenses to plane P_3 . In plane P_3 the magnitude of the energy distribution is recorded. Figure 4 shows the energy distribution of a signal with five different spatial frequencies.

There are a number of coherent optical filtering operations that can be performed on the 1-D signal. For example, using the same arrangement as in Fig. 1, with the exception of placing a vertical slit on the two masks in the input plane, another type of filtering operation can be realized. Figure 5 shows the results of this filtering operation on two and five spatial frequencies, with and without de filtering in the WD plane. The results depict the magnitude of the modified WD. The modification is the additional Fresnel filtering due to the cylindrical lens. Note the much reduced effect of the sidelobes.

Figure 6 depicts the coherent optical implementation of the SFT. There are a number of window functions discussed in the literature.³⁰ In our experiment we used a uniform gate, approximated by a 45° slit, as the window function. The running filter can also be placed in the Fourier plane to affect the SFT. Placing the running filter in the Fourier optical plane of the 1-D signal may be advantageous if we wish to realize a real-time acoustooptic (AO) implementation of the SFT. Since there are physical limitations on the size of the active area that the present AO transducers will support, the 45° slit will degrade the performance of the SFT. Placing the slit in the Fourier plane removes some of the burden on the AO modulator. The magnitude square of the SFT is the spectrogram. By placing a square-law device in the SFT plane, such as photographic film or a gamma-corrected video camera, the spectrogram of the 1-D signal is generated. Figure 7(a) shows the spectrogram of a double-sided frequency ramp. The spectrogram with an impulsive running filter function has been termed as the local spectra¹⁵ of the 1-D signal. It is possible to generate space delay and spatial frequency delay spectrograms by magnitude squaring the 2-D Fourier transform of the SFT. Figure 7(b) shows the delay spectrogram of the double-sided frequency ramp. For an impulsive running filter function this spectra may be termed as the local Doppler spectra. The local Doppler spectra does not have the many sidelobes characteristic of the global Doppler spectrum. Figure 8 shows the local frequency and the local Doppler frequency spectra of five different spatial frequency quasi-periodic signals. Note the excellent frequency resolution of both local spectra.

V. Summary and Conclusions

A new signal representation has been introduced that allows the simultaneous space and spatial frequency filtering of 1-D signals. The new representation, depending on the particular filter function, can represent the WD, the AF, various filtered WD and quasi-WD functions as well as different spectral representations such as the spectrogram, the local spectra, and the local Doppler spectra. Using the 2-D nature of optical signal

processors all the above GSF can be displayed as well as manipulated. Experimental results, using coherent space-integrating optical processors, have been presented. Other types of optical processor can also be modified to display the GSF. These processors will find applications where the real-time display of time-frequency information is of interest.

This work was supported in part under a grant from the Air Force Office of Scientific Research AFOSR 81-0169 and a contract from the Rome Air Development Center F19628-80-C-0095.

References

1. L. J. Cutrona, E. N. Leith, C. J. Palermo, and L. J. Porcello, *IRE Trans. Inf. Theory* **IT-6**, 386 (1960).
2. L. Cutrona, "Recent Developments in Coherent Optical Technology," in *Optical and Electro-optical Information Processing*, J. T. Tippet *et al.*, Eds. (MIT Press, Cambridge, 1965), pp. 83-123.
3. C. E. Thomas, *Appl. Opt.* **5**, 1782 (1966).
4. W. T. Rhodes and J. M. Florence, *Appl. Opt.* **15**, 3073 (1976).
5. J. W. Goodman, P. Kellman, and E. W. Hansen, *Appl. Opt.* **16**, 733 (1977).
6. R. J. Marks II, J. F. Walkup, M. O. Hagler, and T. F. Krile, *Appl. Opt.* **16**, 739 (1977).
7. D. Casasent and D. Psaltis, *Proc. IEEE* **65**, 77 (1977).
8. J. W. Goodman, "Linear Space Variant Optical Processing," in *Optical Data Processing, Fundamentals*, S. Lee, Ed. (Springer, Berlin, 1980).
9. J. F. Walkup, *Opt. Eng.* **19**, 339 (1980).
10. R. A. K. Sood and D. C. Cooper, *Proc. Inst. Electr. Eng.* **120**, 423 (1973).
11. R. J. Marks II, J. F. Walkup, and T. F. Krile, *Appl. Opt.* **16**, 746, 1777 (1977).
12. B. E. A. Saleh, *Appl. Opt.* **17**, 3408 (1978).
13. E. Wigner, *Phys. Rev.* **40**, 719 (1932).
14. M. J. Bastiaans, *J. Opt. Soc. Am.* **69**, 1710 (1979).
15. H. O. Bartelt, K. H. Brenner, and A. W. Lohmann, *Opt. Commun.* **32**, 32 (1980).
16. L. Cohen, *J. Math. Phys. Cambridge, Mass.* **7**, 781 (1966).
17. A. W. Rihaczek, *Principles of High Resolution Radar* (McGraw-Hill, New York, 1969).
18. N. G. DeBruin, *Nieuwe Arch. Wiskunde* **21**(3), 205 (1973).
19. A. W. Rihaczek, *IEEE Trans. Inf. Theory* **IT-14**, 369 (1968).
20. N. G. deBruin, "Uncertainty Principles in Fourier Analysis," in *Inequalities*, O. Shisha, Ed. (Academic, New York, 1967), pp. 57-71.
21. J. B. Allen and L. R. Rabiner, *Proc. IEEE* **65**, 1558 (1977).
22. T. M. Turpin, *Proc. IEEE* **69**, 79 (1981).
23. P. Kellman, *Proc. Soc. Photo Opt. Instrum. Eng.* **185**, 130 (1979).
24. T. R. Bader, "Coherent Hybrid Optical Processors," *Proc. Soc. Photo Opt. Instrum. Eng.* **232**, 140 (1980).
25. T. C. Lee, J. Reibelholz, P. N. Tamura, and J. Lindquist, *Appl. Opt.* **19**, 895 (1980).
26. D. Casasent and B. V. K. Vijaya Kumar, *Appl. Opt.* **18**, 1673 (1979).
27. P. N. Tamura, J. J. Reibelholz, and T. C. Lee, *Opt. Lett.* **5**, 401 (1980).
28. J. D. Cohen, *Proc. Soc. Photo Opt. Instrum. Eng.* **180**, 225 (1979).
29. M. J. Bastiaans, *Appl. Opt.* **19**, 192 (1980).
30. F. J. Harris, *Proc. IEEE* **66**, 51 (1978).

Random TDMA Access Protocol with Application to Multi-Beam Satellites

M. Kawai

T. N. Saadawi

D. L. Schilling

Nippon T & T Co.
Kanagawa-Ken 238-03
Japan

The City College of New York
Electrical Engineering Dept.
New York, N. Y. 10031

ABSTRACT

In this paper, we analyze a random TDMA access protocol for satellite channels. The analysis is based on obtaining the idle and busy period distribution. The average packet queueing delay in the user's buffer is obtained. A modified random TDMA is suggested to improve the performance. Implementation of the scheme in multi-spot beam satellites is presented and the delay on board the satellite is obtained.

1. INTRODUCTION

In packet switched satellite communications systems, various multiaccess schemes have been proposed in order to accomplish flexible connectivity and efficient share of satellite channels. These channel utilization protocols may be partitioned into three main categories. The first is the conventional protocols, such as frequency division multiplexing access TDMA, and time division multiple access, TDMA. These are effective for heavy traffic, but with a bursty nature, such a fixed allocation of channel capacity is extremely wasteful. For such a situation, random access schemes, like the ALOHA protocol, the Um scheme, and the Collision Resolution Algorithms (Tree Algorithm), are introduced. The third category is the dynamic assignment protocols. These include the polling and reservation schemes.

In this paper we analyze a random TDMA multiple access protocol. The analysis is based on obtaining the busy and idle period distributions similar to the approach used in [1] in the analysis of TDMA. The modified random TDMA is then proposed as an improvement to the random TDMA. Other randomized TDMA protocols can be found, for example in [2] and [3]. Implementation of the scheme for multi-spot beam satellites is considered.

A primary concern in satellite communications is to cover a wide area with a high-gain spot beam antenna. A multiple spot beam antenna has a significant effect in making the use of available frequency bands and satellite resources more efficiently. If we consider a system where an antenna is arranged to form N non-overlapping spot beams, each serving a separate zone. Each beam

comprises an uplink and a downlink at different carrier frequencies, with connections made between up and downlinks through a switch matrix in the satellite, shown conceptually in Fig. 1. The switch matrix interconnects receivers and transmitters according to a pre-prescribed algorithm. For simplicity, complete traffic symmetry is assumed: each zone contains the same number of stations L , which is shown by $0 \leq i, j = 1, \dots, N$ and $j = 1, \dots, L$. Each station transmits packets to the same number of stations N , and all interstation packet rates are equal. The arrival process is Poisson and all messages consist of a single packet. All buffer capacities are infinite.

In [4], the authors compare different protocols for accessing a multi-beam satellite. For example, in the Fixed Assignment TDMA, the frame structure is shown in Fig. 2, each earth station directs its traffic to the desired destination zone by transmitting into the preassigned slot during which the switch is set for that destination. The switching sequence in an on-board switch matrix follows a preassigned schedule, possibly modified at intervals by control signals from the ground. Also slotted ALOHA/TDMA systems with multiple uplinks have been investigated for broadcasting satellites in [5], [6], [7] and [8].

In section 2, we describe the random TDMA, and then analyze it in section 3. In section 4 we discuss the implementation of the scheme in multi-spot beam satellites.

2. The Random TDMA Protocol

Here, we will describe the Random TDMA protocol for a global beam covering the entire area to be serviced.

We are considering M identical users. The M users are divided into N groups. Each group consists of L users (i.e., M/N). Each user has an infinite buffer. Messages arrivals at each user follow the statistics of a Poisson process with rate λ . Each message is assumed to consist of a single packet. Channel time is slotted and all users are synchronized to the channel time. The slot length is equal to the duration of one packet. Channel slots are grouped in frames each of which consists of N consecutive slots.

Each group is assigned a fixed favored slot in the frame, as in the case of dedicated slots of

This work was partially supported by AFOSR under grant AFOSR 81-0169

ordinary TDMA, and no two groups have the same favored slot. Each user (or earth station) within a group has the same favored slot.

When a user has a packet ready for transmission (i.e., the moment a packet arrives at the Head-of-the-Line (HOL) position in the buffer), he looks at the next N slots and schedules transmission in the favored slot of his group with probability a and in any other particular slot within the next N slots with probability b . Thus:

$$a + b(N-1) = 1 \quad (1)$$

Let us explain this process using a roulette, which has the numbers 1 to N . One of these numbers appears with probability a , the others appear with probability b . During busy period when there are some packets in the buffer, as soon as a packet at HOL is transmitted, the roulette is turned and hence the HOL packet transmission is scheduled according to the result of the roulette. During idle periods when there is no packets in the buffer, as soon as a packet arrives at the buffer, the roulette schedules the packet transmission, Fig. 3.

If a packet encounters a collision, which is acknowledged after a round-trip time delay, it joins the buffer at the End-of-the-Line (EOL) position in the same way as a packet newly generated. The queuing model with feedback is shown in Fig. 4. The message arrival rate including collided packets is denoted by λ_G .

3. Analysis

Let D_1 be the expected packet delay due to the user's buffer. The expected packet delay in the system is then given by

$$D = 2P + D_1 + E(D_1 + 2P) \quad (2)$$

where

E = the average number of retransmissions
 $2P$ = round-trip delay.

3.1 The Average Packet Waiting Time

To find the value of D_1 , we use an approach similar to the one used by Lam [LAM 77] in his analysis of TDMA. We consider two kinds of service time distribution function $B(x)$ and $B^*(x)$ in a single-server queue with Poisson arrivals at λ_G packets per second $B(x)$ is the distribution of a packet which initiates a busy period with first and second moments b_1 and b_2 . All subsequent packets in the same busy period have service times drawn independently from the distribution $B^*(x)$ with first and second moments b_1^* and b_2^* . Let N_t be the number of packets in the system (both in queue and in service) at time t . We now define the transforms:

$$P(z) = \sum_{n=0}^{\infty} z^n P_n \quad ; \quad P_n = \lim_{t \rightarrow \infty} \text{Prob}[N_t = n] \quad (3)$$

$$\hat{B}(s) = \int_0^{\infty} e^{-sx} dB(x) \quad (4)$$

$$\hat{B}^*(s) = \int_0^{\infty} e^{-sx} dB^*(x) \quad (5)$$

Then in a proof similar to Lam's proof, D_1 is given by

$$D_1 = \frac{b_1}{1 - \lambda_G(b_1 - b_1^*)} + \frac{\lambda_G(b_2 - b_2^*)}{2[1 - \lambda_G(b_1 - b_1^*)]} + \frac{\lambda_G b_2}{2(1 - \lambda_G b_1^*)} \quad (6)$$

After some manipulation shown in Appendix A we obtain

$$B^*(s) = (a-b)^2 e^{-sT} + [2ab + (N-2)b^2] \sum_{i=1}^N e^{-\frac{i}{N}sT} \quad (7)$$

$$B^*(s) = (a-b)^2 \int_0^T e^{-s(y+\frac{T}{N})} dF(y) +$$

$$[2ab + (N-2)b^2] \sum_{i=1}^N e^{-\frac{i}{N}sT} \frac{\lambda_G}{\lambda_G s} \cdot \frac{e^{-\frac{sT}{N}} - e^{-\frac{\lambda_G T}{N}}}{1 - e^{-\frac{\lambda_G T}{N}}} \quad (8)$$

$$F(y) = \frac{e^{-\lambda_G(T-\frac{T}{N})} (e^{\lambda_G y} - 1)}{1 - e^{-\lambda_G T}} + u(y - T + \frac{T}{N})$$

$$= [1 - e^{-\lambda_G(T-\frac{T}{N}-y)}] \quad 0 \leq y \leq T \quad (9)$$

where

$$u(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (10)$$

From eqs. (6), (7), (8) we can easily obtain the first and second moments of packet service times, b_1 , b_2 , b_1^* , b_2^* .

3.2 The Average Number of Retransmissions E

The average number of retransmissions required per packet generated is related under equilibrium conditions to S , the system throughput and G , the channel traffic:

$$1 + E = \frac{G}{S} \quad (11)$$

where, $s = \lambda_G T$ and $G = \lambda_G T$. Here a_1 , the probability that a packet is transmitted in a favored slot and b_1 , the probability that a packet is transmitted in a non-favored slot, are given by $a_1 = a\lambda_G T$ and $b_1 = b\lambda_G T$. Defining P_s to be the probability that a packet is transmitted successfully and P_c to be the probability that a packet collides with one or more other packets, and assuming independence, average numbers of retransmission attempts is given by

$$E = \sum_{i=0}^{\infty} P_s P_c^i = \frac{P_s P_c}{(1 - P_c)^2} \quad (12)$$

where

$$P_s = \text{Prob}[A \text{ successful transmission in a favored slot}] + \text{Prob}[A \text{ successful transmission in non-favored slots}]$$

$$= a P_{10} + b P_{20} \quad (13)$$

$$P_f = \text{Prob [collisions in a favored slot]} + \text{Prob [collisions in non-favored slots]} \\ = \sum_{i=1}^{M-1} P_{1i} + (M-1) \sum_{i=1}^{M-1} P_{2i} \quad (14)$$

P_{1i} = Prob [i packets in (L-1) favored and (M-L) non-favored packets are transmitted simultaneously].

$$= \sum_{j=0}^{i-1} \binom{L-1}{j} a_1^j (1-a_1)^{L-1-j} \binom{M-L}{M-L-i-j} b_1^{M-L-i-j} (1-b_1)^{M-L-i-j}$$

$$i_1 = \min (1, L-1) \quad (15)$$

P_{2i} = Prob [i packets in L favored and (M-L-1) non-favored packets are transmitted simultaneously]

$$= \sum_{j=0}^{i-2} \binom{L-1}{j} a_1^j (1-a_1)^{L-1-j} \binom{M-L-1}{M-L-1-i-j} b_1^{M-L-1-i-j} (1-b_1)^{M-L-1-i-j}$$

$$i_2 = \min (1, L)$$

$$\text{and } N_K = \binom{N}{K} \quad (16)$$

4. Multi-Beam Satellite

Here, we consider the implementation of random TDMA in multi-spot beam satellite. The system model is the same as in Fig. 1.

Each user has N buffers each of which corresponds to the different destination. Each buffer employs the above-mentioned random access scheme. The main procedure peculiar to the multibeam system in the assignment of the favored slot. The favored slot is assigned into the n^{th} slot in a frame for the packets which are transmitted from uplink A_1 to downlink A_j , where

$$h = \begin{cases} 1+j-1 & : 1+j \leq N+1 \\ 1+j-1-N & : 1+j > N+1 \end{cases} \quad (17)$$

When a packet is acknowledged to encounter a collision after a certain delay, it is retransmitted at the EDL in the user's buffer as described in Sec. 2. Only non-collided packets are accepted into the processing satellite with buffering capability and transmitted to the appropriate downlink on the basis of TRFA.

In this case the expected packet delay is given by

$$D_M = 2F + D_1 + D_2 + E(D_1 + 2F) \quad (18)$$

where D_1 = the expected packet delay at the satellite. The difference from Eq. (3) is the term D_2 due to the queuing delay in the satellite. The expected number of packets in a satellite buffer for a specific destination is given by, (see Appendix B).

$$\bar{N} = \frac{a_2^2 (N-1) b_2}{[1-a_1^2 (N-1) b_2]} \left[1 - \frac{a_2^2 (N-1) b_2}{2} \left(1 + \frac{a_2^2 (N-1) b_2^2}{[a_2^2 (N-1) b_2]} \right) \right]$$

$$a_2 = a_1^2 T_L, \quad b_2 = b_1^2 T_L \quad (19)$$

An application of Little's formula gives the expected packet delay $D_2 = N \bar{N} / G$.

Figure 5 shows the analytical and simulation results for the average packet queuing delay in

the earth station versus λ_1 , the channel traffic rate. Both results coincide with each other. It is also noted that the effect of $a_1 b_1$ is small at lower rates and quite large at higher channel traffic rates due to increased collision.

Figure 6 shows the analytical and simulation results for the total average packet delay D as given in Eq. (3), versus throughput. Note that for higher λ_1 (i.e. close to TRFA) we obtain better performance.

The results can be improved at lower throughput by making use of the fact that the idle periods, when no packets are in the user's buffer, are dominant at lower throughput and busy periods are dominant at higher throughput. The modified random TRFA is described as follows;

During idle periods, the arrived packets are assumed to be transmitted before the end of the frame, which has the favored slot as the last slot. Namely, when a packet arrives at a slot in a frame consisting of N slots, it is transmitted between $(n+1)^{\text{th}}$ and n^{th} , where the location of the favored slot is the n^{th} slot, with equal probability. The simulation results for the average packet queuing delay are shown in Fig. 5. It shows improvements at lower throughput compared to TRFA and random TRFA. Figure 6 shows the simulation results for the delay throughput characteristics, which also shows the advantage of the modified random TRFA at lower throughput.

Appendix A

In this appendix we derive Eqs. (7) to (9). First of all we consider the relation between two consecutive packet transmissions. Let i (1 to N) and j (1 to N) represent the slot addresses of these consecutive packets. When we assume a favored slot address with #1, non-favored slot addresses with #2 to #N, we obtain the probability

$$\text{Prob} (i, j) = \begin{cases} a^2 & ; i = 1 \\ ab & ; i = 1, j \neq 1 \text{ or } i \neq 1, j = 1 \\ b^2 & ; i \neq 1, j \neq 1 \end{cases} \quad (A1)$$

During busy periods, the probability that the service time is kT_N is given by

$$P_k = \text{Prob} (i, j) \cdot w(kT_N), \quad i, k, N \quad (A2)$$

where

$$w(x) = \begin{cases} x/N & ; x \leq 0 \\ x & ; 0 < x \leq N \\ x-N & ; N < x \end{cases} \quad (A3)$$

Therefore, we obtain

$$\frac{dD(x)}{dx} = \sum_{k=1}^N \sum_{j=1}^N \text{Prob} \{i, j = w(i+k)\} \delta(x - \frac{kT}{N}) \quad (A4)$$

where

$\delta(x)$ is a δ -function.

Equations (4), (A1) and (A4) give (6).

Now we proceed to deal with the case that an idle period is included between two consecutive packet transmissions. In this case, i and j represent the slot addresses for the last transmission of the preceding busy period and the first transmission of the idle period. Here we introduce a variable y , which is defined to be the period between the first arrival and the corresponding location in the current frame of the last departure, as shown in Fig. 4. Since the variable has the same meaning as that defined in [LBN 78], the distribution function is given by Eq. (9). Here, when y is in the region

$$\frac{h-1}{N} T \leq y < \frac{h}{N} T ; 1 \leq h \leq N,$$

the probability that the service time is $y + \frac{kT}{N}$ is given by

$$\sum_{i=1}^N \text{Prob} \{i, j = w(i+k-1)\}$$

$$; 1 \leq h+k-1 \leq N$$

Therefore we obtain

$$\frac{dD(x)}{dx} = \sum_{k=2-h}^{N+1-h} \sum_{i=1}^N \text{Prob} \{i, j = w(i+k-1)\} \frac{dF(y)}{dy} \bigg|_{y=x-\frac{kT}{N}}$$

$$\frac{h+k-1}{N} T \leq x < \frac{h+k}{N} T \quad (A5)$$

Equations (5), (A1) and (A5) give eqn. (7)

Appendix B

In this appendix we derive Eq. (22). Here we think about a concentrator with N terminals. The concentrator output transmission line rate is one data unit per Δ sec. Each terminal is synchronized to this output rate. One of these terminals feeds in either one data unit with probability a every Δ sec, the others with probability b , every Δ sec. In this case the probability that k packets arrive in the concentrator simultaneously is given by

$$v(k) = \begin{cases} (1-a_2)(1-b_2)^{N-1} & ; k=0 \\ (1-a_2)^{N-1} C_k b_2^k (1-b_2)^{N-1-k} + \\ a_2^{N-1} C_{k-1} b_2^{k-1} (1-b_2)^{N-k} & ; 1 \leq k \leq N \\ a_2 b_2^{N-1} & ; k=N \end{cases} \quad (B1)$$

The Z-transform of Eq. (B1) is

$$V(z) = \sum_{k=0}^N v(k) z^k \quad (B2)$$

$$= (a_2 z + 1 - a_2) (a_2 z + 1 - b_2)^{N-1}$$

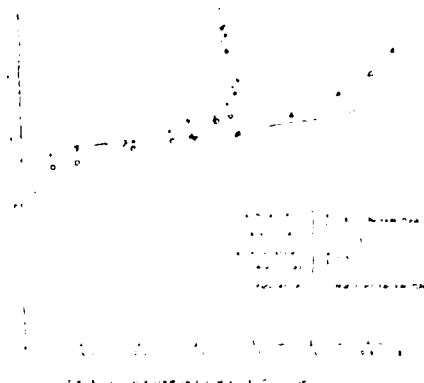
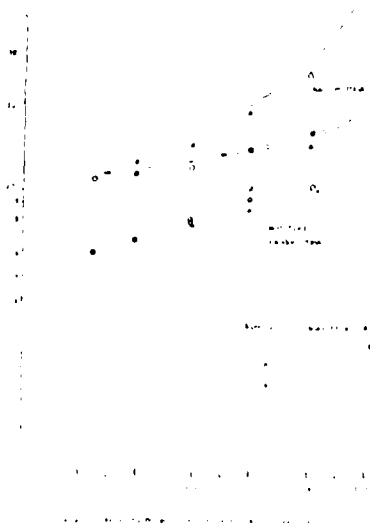
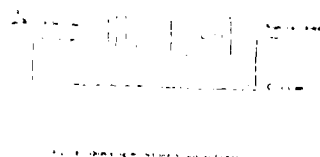
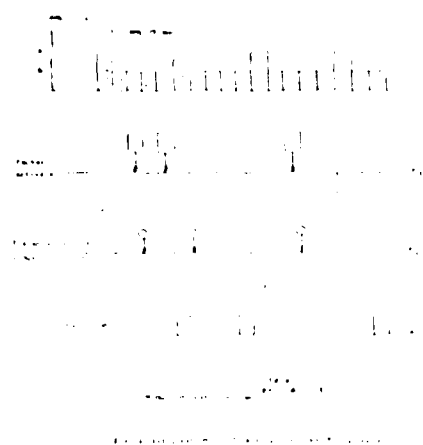
Substituting (B2) into the Pollaczek-Khinchine transform Eq. [KCHW 77], we have the current-generating function

$$Q(z) = (1-\rho) \frac{(z-1)V(z)}{z-V(z)} \quad (B3)$$

$$\rho = \frac{dQ(z)}{dz} \bigg|_{z=1} = a_2 + (N-1)b_2 \quad (B4)$$

Therefore, the expected number of messages in a satellite buffer for a specific destination is given by $N = \frac{dQ(z)}{dz} \bigg|_{z=1}$ which gives eqn. (19).





References

- [IEEE 78] IEEE, J.K. and L.H. Green, "Packet Switching in a Processing Satellite", Proc. IEEE, Vol. 66, pp. 100-102, Jan. 1978.
- [CAVE 79] Cave, R.E., "ANITA/TM Systems with Multiple Downlink Capacity," IEEE Trans. Comm. vol. COM-27, pp. 537-541, March 79.
- [COMR 80] Gatre J. and T.E. Stern, "A Comparison of Multiple Access Protocols for Packet Switching in Satellite Switched Multiservice Systems", ICC Conf. pp. 87, pp. 53-41-6, June 1980.
- [IEEE 80] J. Green, A. and O. B. S. "Analysis of a Hybrid Access Scheme for Buffered Unrestricted Plastic Time Division", ICC 1980, pp. 71-6.1-71-6.8.
- [IEEEIN 78] Fleishman, L. and Y. Yehani, "An Optimal Adaptive Scheme for Multiple Access Broadcast Communications", ICC June 78, pp. 7.2.1-7.2.5.
- [IEEE 77] Lee, S. B. "Timing Analysis of a Time Division Multiple Access Channel", IEEE Trans. Tex. 77.
- [IEEE 77] Schwartz, M., "Computer-Communication Network Design and Analysis" Prentice Hall Inc.
- [IEEE 80] Sali, T.M. Maydara and T. B. S. "Optical Backscatter Assignment on up and downlinks of Satellite with Buffer Capacity," IEEE Trans. Comm. Vol. COM-28, pp. 1609-1616, Oct. 1980.

input is a phase mask. The transilluminated mask is in turn imaged using two spherical lenses to plane P_2 . In plane P_2 , the magnitude of the energy distribution is recorded. Figure 4 shows the energy distribution of a signal with five different spatial frequencies.

There are a number of coherent optical filtering operations that can be performed on the 1-D signal. For example, using the same arrangement as in Fig. 1, with the exception of placing a vertical slit on the two masks in the input plane, another type of filtering operation can be realized. Figure 5 shows the results of this filtering operation on two and five spatial frequencies, with and without dc filtering in the WD plane. The results depict the magnitude of the modified WD. The modification is the additional Fresnel filtering due to the cylindrical lens. Note the much reduced effect of the sidelobes.

Figure 6 depicts the coherent optical implementation of the SFT. There are a number of window functions discussed in the literature.³⁰ In our experiment we used a uniform gate, approximated by a 45° slit, as the window function. The running filter can also be placed in the Fourier plane to affect the SFT. Placing the running filter in the Fourier optical plane of the 1-D signal may be advantageous if we wish to realize a real time acoustooptic (AO) implementation of the SFT. Since there are physical limitations on the size of the active area that the present AO transducers will support, the 45° slit will degrade the performance of the SFT. Placing the slit in the Fourier plane removes some of the burden on the AO modulator. The magnitude square of the SFT is the spectrogram. By placing a square law device in the SFT plane, such as photographic film or a *gamma* corrected video camera, the spectrogram of the 1-D signal is generated. Figure 7(a) shows the spectrogram of a double-sided frequency ramp. The spectrogram with an impulsive running filter function has been termed as the local spectra¹⁵ of the 1-D signal. It is possible to generate space delay and spatial frequency delay spectrograms by magnitude squaring the 2-D Fourier transform of the SFT. Figure 7(b) shows the delay spectrogram of the double sided frequency ramp. For an impulsive running filter function this spectra may be termed as the local Doppler spectra. The local Doppler spectra does not have the many sidelobes characteristic of the global Doppler spectrum. Figure 8 shows the local frequency and the local Doppler frequency spectra of five different spatial frequency quasi-periodic signals. Note the excellent frequency resolution of both local spectra.

V. Summary and Conclusions

A new signal representation has been introduced that allows the simultaneous space and spatial frequency filtering of 1-D signals. The new representation, depending on the particular filter function, can represent the WD, the AF, various filtered WD and quasi-WD functions as well as different spectral representations such as the spectrogram, the local spectra, and the local Doppler spectra. Using the 2-D nature of optical signal

processors all the above GSF can be displayed as well as manipulated. Experimental results, using coherent space integrating optical processors, have been presented. Other types of optical processor can also be modified to display the GSF. These processors will find applications where the real-time display of time-frequency information is of interest.

This work was supported in part under a grant from the Air Force Office of Scientific Research AFOSR 81-0169 and a contract from the Rome Air Development Center F19628-80-C-0095.

References

1. L. J. Cutrona, F. N. Leith, C. J. Palermo, and L. J. Porcello, *IRE Trans. Inf. Theory* **14**, 80 (1966).
2. L. Cutrona, "Recent Developments in Coherent Optical Technology," in *Optical and Electro-Optical Information Processing*, J. T. Tappett et al., Eds. (MIT Press, Cambridge, 1965), pp. 83-123.
3. C. E. Thomas, *Appl. Opt.* **5**, 1782 (1966).
4. W. F. Rhodes and J. M. Florence, *Appl. Opt.* **15**, 3073 (1976).
5. J. W. Goodman, P. Kellman, and E. W. Hansen, *Appl. Opt.* **16**, 1713 (1977).
6. R. J. Marks II, J. F. Walkup, M. O. Hügler, and T. F. Krile, *Appl. Opt.* **16**, 139 (1977).
7. D. Casasent and D. Psaltis, *Proc. IEEE* **65**, 77 (1977).
8. J. W. Goodman, "Linear Space-Variant Optical Processing," in *Optical Data Processing, Fundamentals*, S. Lee, Ed. (Springer, Berlin, 1980).
9. J. F. Walkup, *Opt. Eng.* **19**, 339 (1980).
10. R. A. K. Sadi and D. C. Cooper, *Proc. Inst. Electr. Eng.* **120**, 423 (1973).
11. R. J. Marks II, J. F. Walkup, and T. F. Krile, *Appl. Opt.* **16**, 746, 1777 (1977).
12. B. E. A. Saleh, *Appl. Opt.* **17**, 3108 (1978).
13. E. Wigner, *Phys. Rev.* **40**, 719 (1935).
14. M. J. Bastiaans, *J. Opt. Soc. Am.* **69**, 1710 (1979).
15. H. O. Barchi, K. H. Brenner, and A. W. Lohmann, *Opt. Commun.* **32**, 32 (1980).
16. L. Cohen, *J. Math. Phys. Cambridge, Mass.* **7**, 781 (1966).
17. A. W. Ribacek, *Principles of High Resolution Radar* (McGraw-Hill, New York, 1969).
18. N. G. DeBruin, *Nieuwe Arch. Wiskunde* **21**(3), 205 (1973).
19. A. W. Ribacek, *IEEE Trans. Inf. Theory* **14**, 369 (1968).
20. N. G. deBruin, "Uncertainty Principles in Fourier Analysis," in *Inequalities*, O. Shisha, Ed. (Academic, New York, 1967), pp. 57-71.
21. J. B. Allen and L. R. Rabiner, *Proc. IEEE* **65**, 1558 (1977).
22. T. M. Turpin, *Proc. IEEE* **69**, 79 (1981).
23. P. Kellman, *Proc. Soc. Photo Opt. Instrum. Eng.* **185**, 130 (1979).
24. T. R. Bader, "Coherent Hybrid Optical Processors," *Proc. Soc. Photo Opt. Instrum. Eng.* **232**, 140 (1980).
25. T. C. Lee, J. Rebholz, P. N. Tamura, and J. Lindquist, *Appl. Opt.* **19**, 895 (1980).
26. D. Casasent and B. V. K. Vajaya Kumar, *Appl. Opt.* **18**, 1673 (1979).
27. P. N. Tamura, J. J. Rebholz, and T. C. Lee, *Opt. Lett.* **5**, 401 (1980).
28. J. D. Cohen, *Proc. Soc. Photo Opt. Instrum. Eng.* **180**, 225 (1979).
29. M. J. Bastiaans, *Appl. Opt.* **19**, 192 (1980).
30. F. J. Harris, *Proc. IEEE* **66**, 51 (1978).

Random TDMA Access Protocol with Application to Multi-Beam Satellites

M. Elwan

T. N. Saadawi

D. L. Schilling

Support Systems
Research and Development
Japan

The City College of New York
Electrical Engineering Dept.
New York, N. Y. 10031

ABSTRACT

In this paper, we analyze a random TDMA access protocol for satellite channels. The analysis is based on obtaining the idle and busy period distribution. The average packet queuing delay in the user's buffer is obtained. A modified random TDMA is suggested to improve the performance. Implementation of the scheme in multi-spot beam satellites is presented and the delay on board the satellite is obtained.

1. INTRODUCTION

In packet switched satellite communications systems, various multiaccess schemes have been proposed in order to accomplish flexible connectivity and efficient share of satellite channels. These channel utilization protocols may be partitioned into three main categories. The first is the conventional protocols, such as frequency division multiplexing access TDMA, and time division multiple access, TDMA. These are effective for heavy traffic, but with a bursty nature, such a fixed allocation of channel capacity is extremely wasteful. For such a situation, random access schemes, like the ALOHA protocol, the CSMA scheme, and the Collision Resolution Algorithms (Tree Algorithm), are introduced. The third category is the dynamic assignment protocols. These include the polling and reservation schemes.

In this paper we analyze a random TDMA multiple access protocol. The analysis is based on obtaining the busy and idle period distributions similar to the approach used in [LAW 78] in the analysis of TDMA. The modified random TDMA is then proposed as an improvement to the random TDMA. Other randomized TDMA protocols can be found, for example in [FLEI 78] and [EFER 80]. Implementation of the scheme for multi-spot beam satellites is considered.

A primary concern in satellite communications is to cover a wide area with a high-gain spot beam antenna. A multiple spot beam antenna has a significant effect in making the use of available frequency bands and satellite resources more efficiently. If we consider a system where antenna is arranged to form N non-overlapping spot beams, each serving a separate zone. Each beam

surprises an uplink and a downlink at different carrier frequencies, with connections made between up and downlinks through a switch matrix in the satellite, shown conceptually in Fig. 1. The switch matrix interconnects receivers and transmitters according to a pre-prescribed algorithm. For simplicity, complete traffic symmetry is assumed: each zone contains the same number of stations L , which is chosen to be $1, 2, \dots, N$ and $1, 2, \dots, L$. Each station transmits packets to the same number of stations L and all interstation packet rates are equal. The arrival process is Poisson and all messages consist of a single packet. All buffer capacities are infinite.

In [GWER 80], the authors compare different protocols for accessing a multi-beam satellite. For example, in the Fixed Assignment TDMA, the frame structure is shown in Fig. 2, each earth station directs its traffic to the desired destination zone by transmitting into the preassigned slot during which the switch is set for that destination. The switching sequence in an on-board switch matrix follows a prearranged schedule, possibly modified at infrequent intervals by control signals from the ground. Also slotted ALOHA/TDMA systems with multiple uplinks have been investigated for broadcasting satellites in [SUDA 80], [LEWE 79] and [BIRN 78].

In section 2, we describe the random TDMA, and then analyze it in section 3. In section 4 we discuss the implementation of the scheme in multi-spot beam satellites.

2. The Random TDMA Protocol

Here, we will describe the Random TDMA protocol for a global beam covering the entire area to be serviced.

We are considering M identical users. The M users are divided into N groups. Each group consists of L users ($1 \leq L, M \leq N$). Each user has an infinite buffer. Messages arrivals at each user follow the statistics of a Poisson process with rate λ . Each message is assumed to consist of a single packet. Channel time is slotted and all users are synchronized to the channel time. The slot length is equal to the duration of one packet. Channel slots are grouped in frames each of which consists of N consecutive slots.

Each group is assigned a fixed favored slot in the frame, as in the case of dedicated slots of

This work was partially supported by AFOSR under grant AFOSR 81-0169

$$P_f = \text{Prob [collisions in a favored slot]} \\ + \text{Prob [collisions in non-favored slots]} \\ = a \sum_{i=1}^{M-1} P_{1i} + (3-1) b \sum_{i=1}^{M-1} P_{2i} \quad (14)$$

P_{1i} = Prob [i packets in (L-1) favored and (M-1) non-favored packets are transmitted simultaneously].

$$= \frac{1}{j=0} \frac{L-1}{L-1} a_1^j (1-a_1)^{L-1-j} \frac{M-1}{M-1} b_1^{i-j} (1-b_1)^{M-1-i+j} \quad (15)$$

P_{2i} = Prob [i packets in L favored and (M-1) non-favored packets are transmitted simultaneously]

$$= \frac{1}{j=0} \frac{L-1}{L-1} a_1^j (1-a_1)^{L-1-j} \frac{M-1}{M-1} b_1^{i-j} (1-b_1)^{M-1-i+j} \quad (16)$$

$$\text{and } N_K = \binom{N}{K} \quad (17)$$

4. Multi-Beam Satellite

Here, we consider the implementation of random TDMA multi-spot beam satellite. The system model is the same as in Fig. 1.

Each user has N buffers each of which corresponds to the different destination. Each buffer obeys the conventional random access scheme. The main procedure peculiar to the multi-beam system is the assignment of the favored slot. The favored slot is assigned into the k_1 slot in a frame for the packets which are transmitted from uplink A_1 to downlink A_j , where

$$h = \begin{cases} (1+j)-1 & : (1+j) \leq N \\ (1+j)-N & : (1+j) > N \end{cases} \quad (17)$$

When a packet is acknowledged to encounter a collision after a certain delay, it is retransmitted at the DJL in the user's buffer as described in Sec. 2. Only non-collided packets are accepted into the processing satellite with buffering capability and transmitted to the appropriate downlink on the basis of TMA.

In this case the expected packet delay is given by

$$D_M = 2P + D_1 + D_2 + E(D_1 + 2P) \quad (18)$$

where D_1 is the expected packet delay at the satellite. The difference from Eq. (3) is the term D_2 due to the queuing delay in the satellite. The expected number of packets in a satellite buffer for a specific destination is given by, (see Appendix B),

$$\bar{N} = \frac{a_2^2 (N-1) b_2}{[1-a_1^2 (N-1) b_2]} \left[1 - \frac{a_2^2 (N-1) b_2}{2} \left(1 + \frac{a_2^2 (N-1) b_2}{[a_2^2 (N-1) b_2]} \right) \right]$$

$$D_2 = a_2^2 H_2, H_2 = E[N_2] \quad (19)$$

The application of Little's formula gives the expected packet delay $D_2 = E[N_2] H_2$.

Figure 5 shows the analytical and simulation results for the average packet queuing delay in

the earth station versus b_1 , the channel traffic rate. Both results coincide with each other. It is also noted that the effect of a delay is small at lower rates and quite large at higher channel traffic rates due to increased collision.

Figure 6 shows the analytical and simulation results for the total average packet delay D as given in eq. (18), versus throughput. Note that for higher b_1 (i.e. close to TMA) we obtain better performance.

The results can be improved at lower throughput by taking into account the fact that the idle period, when no packets are in the user's buffer, are dominant at lower throughput and long periods are dominant at higher throughput. The modified multi-TMA is described as follows:

During idle periods, the arrived packets are assumed to be transmitted at the end of the frame, which has the favored slot as the last slot. Namely, when a packet arrives at a slot in a frame consisting of N slots, it is transmitted between $(N-1)th$ and Nth , where the location of the favored slot is the Nth slot, with equal probability. The simulation results for the average packet queuing delay are shown in Fig. 5. It is seen that the results at lower throughput are equal to TMA and random TMA. Figure 6 shows the simulation results for the delay throughput characteristics, which also shows the advantage of the modified random TMA at lower throughput.

APPENDIX A

In this appendix we derive Eqs. (7) to (9). First of all we consider the relation between two consecutive packet transmissions. Let i ($1 \leq i \leq N$) and j ($1 \leq j \leq N$) represent the slot addresses of the i consecutive packets. When we assume a favored slot address with #1, non-favored slot addresses with #2 to #N, we obtain the probability

$$\text{Prob} [i, j] = \begin{cases} a^2 & , i=j=1 \\ ab & , i=1, j \neq 1 \text{ or } i \neq 1, j=1 \\ b^2 & , i \neq 1, j \neq 1 \end{cases} \quad (A1)$$

During busy periods, the probability that the service time is KT/N is given by

$$\text{Prob} [i, j] = w(i, k), \quad i, k \leq N \quad (A2)$$

where

$$w(x) = \begin{cases} x \cdot N & : x \leq 0 \\ x & : 0 < x \leq N \\ x - N & : N < x \end{cases} \quad (A3)$$

Therefore, we obtain

$$\frac{dD(x)}{dx} = \sum_{k=1}^N \sum_{i=1}^N \text{Prob} \{i, j\} w(i, k) \delta(x - \frac{kT}{N}) \quad (A4)$$

where

$\delta(x)$ is a δ -function.

Equations (4), (A1) and (A4) give (6).

Now we proceed to deal with the case that an idle period is included between two consecutive packet transmissions. In this case, i and j represent the slot addresses for the last transmission of the preceding busy period and the first transmission of the idle period. Here we introduce a variable y , which is defined to be the period between the first arrival and the corresponding location in the current frame of the last departure, as shown in Fig. 4. Since the variable has the same meaning as that defined in [12, 7b], the distribution function is given by Eq. (9). Here, when y is in the region

$$\frac{h-1}{N} T \leq y < \frac{h}{N} T, \quad 1 \leq h \leq N,$$

the probability that the service time is $y + \frac{kT}{N}$ is given by

$$\sum_{i=1}^N \text{Prob} \{i, j\} = w(i, k, h)$$

$$; \quad 1 \leq h \leq N$$

Therefore we obtain

$$\frac{dD(x)}{dx} = \sum_{k=2-h}^{N-1-h} \sum_{i=1}^N \text{Prob} \{i, j\} w(i, k, h) \frac{dD(y)}{dy} \Big|_{y=x-\frac{kT}{N}}$$

Equations (5), (A1) and (A4) give eqn. (7).

Appendix B

In this appendix we derive Eq. (22). Here we think about a concentrator with N terminals. The concentrator output transmission line rate is one data unit per Δ sec. Each terminal is synchronized to this output rate. One of these terminals feeds in either one data unit with probability a every Δ sec, the others with probability b , every Δ sec. In this case the probability that k packets arrive in the concentrator simultaneously is given by

$$v(k) = \begin{cases} (1-a_2)(1-b_2)^{N-1} & ; k=0 \\ (1-a_2)^{N-1} C_k b_2^k (1-b_2)^{N-1-k} + \\ a_2^{N-1} C_{k-1} b_2^{k-1} (1-b_2)^{N-k} & ; 1 \leq k \leq N \\ a_2 b_2^{N-1} & ; k=N \end{cases} \quad (B1)$$

The Z-transform of Eq. (B1) is

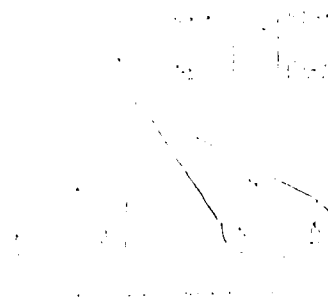
$$V(z) = \sum_{k=0}^N v(k) z^k = (a_2 z + 1 - a_2) (b_2 z + 1 - b_2)^{N-1} \quad (B2)$$

Substituting (B2) into the Pollaczek-Khinchine transform Eq. [KTM 77], we have the moment-generating function

$$Q(z) = (1-\rho) \frac{(z-1) V(z)}{z-V(z)} \quad (B3)$$

$$\rho = \frac{Q'(1)}{Q(1)} \Big|_{z=1} = a_2 + (N-1)b_2 \quad (B4)$$

Therefore, the expected number of messages in a satellite buffer for a specific destination is given by $N = a_2'(z)$ which gives eqn. (19).



$$Q'(z) = \frac{dQ(z)}{dz} = (1-\rho) \frac{(z-1) V'(z) + V(z)}{z^2 - V(z)}$$

$$Q'(1) = \frac{dQ(z)}{dz} \Big|_{z=1} = \frac{(1-\rho) V'(1)}{1-V(1)}$$

$$Q'(1) = \frac{(1-\rho) V'(1)}{1-V(1)} = \frac{(1-\rho) V'(1)}{1-V(1)}$$

$$Q'(1) = \frac{(1-\rho) V'(1)}{1-V(1)} = \frac{(1-\rho) V'(1)}{1-V(1)}$$

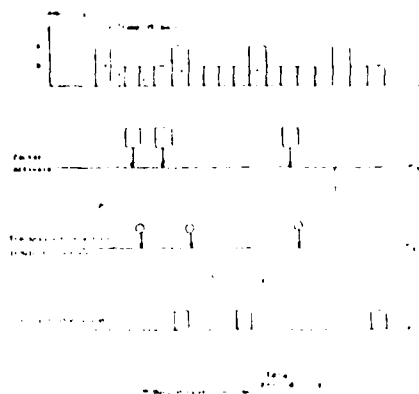


FIG. 1. Sequence of events.



FIG. 2. Sequence of events.

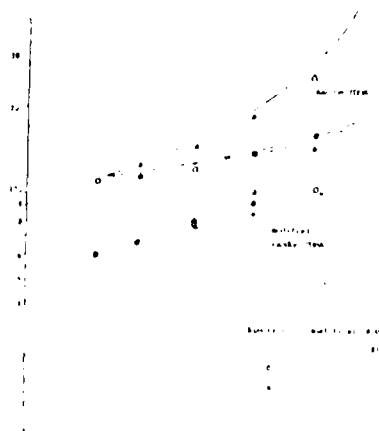
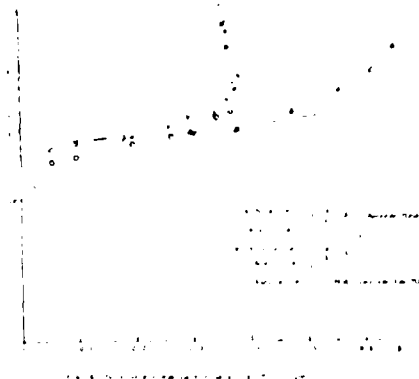
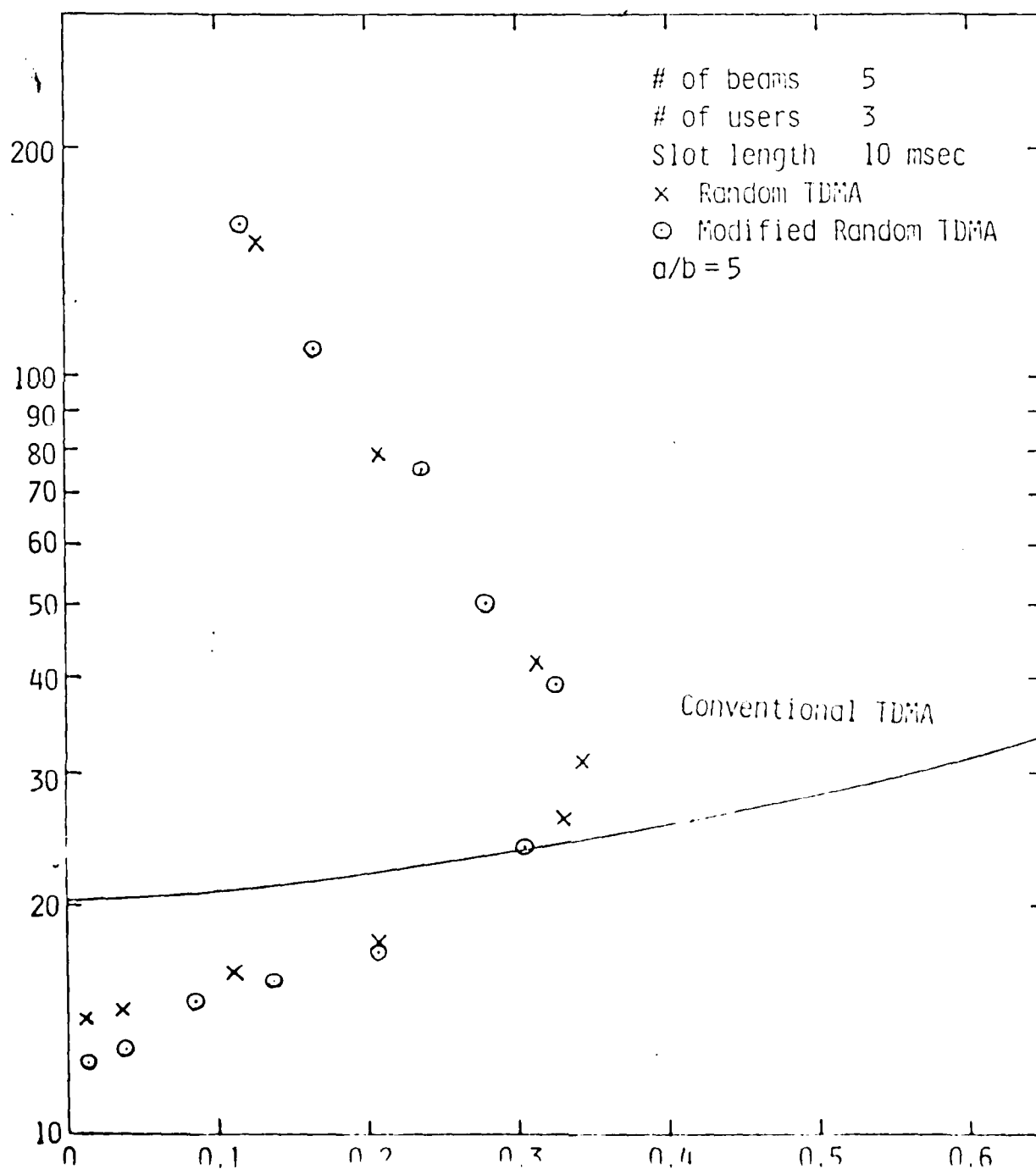


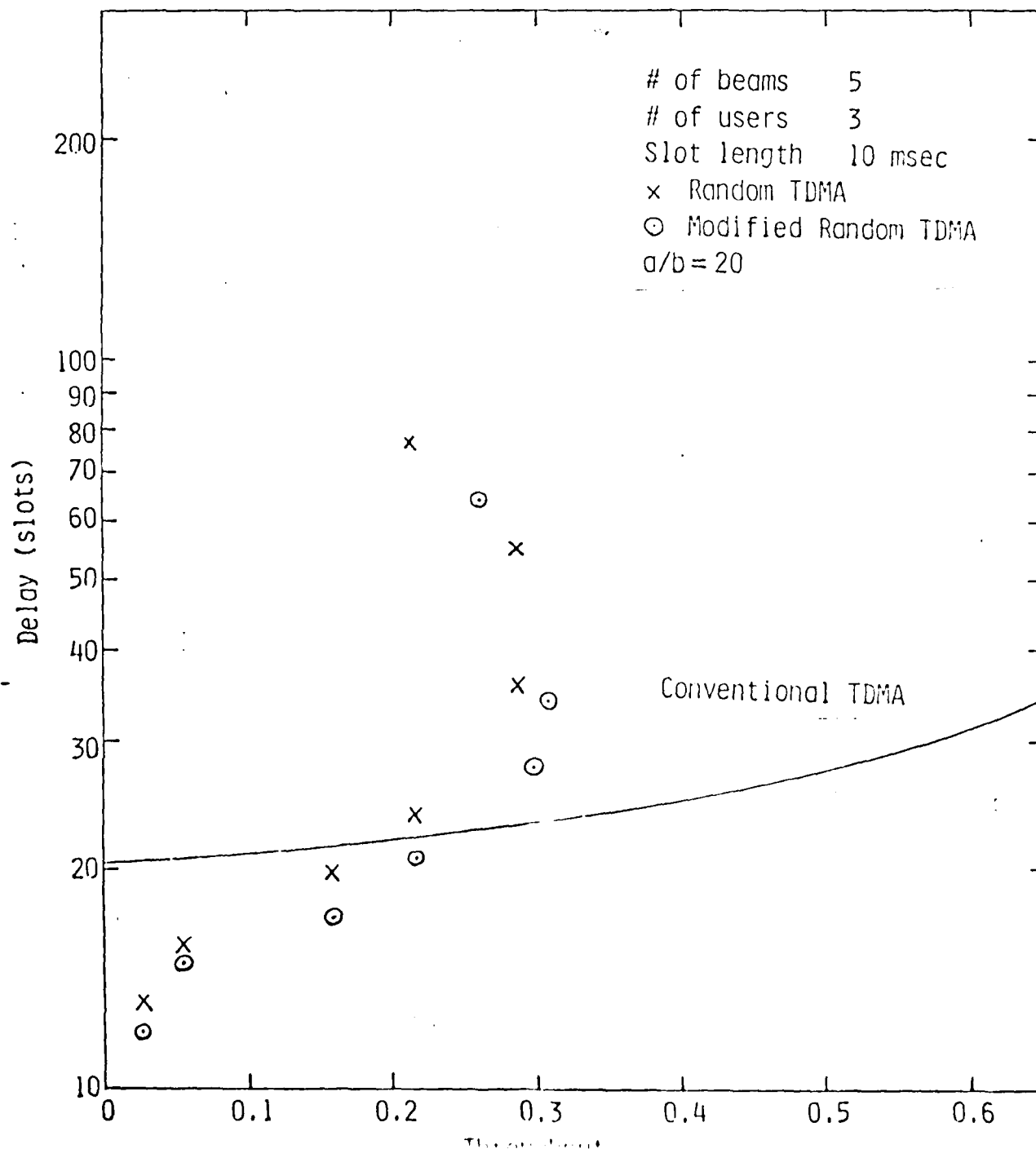
FIG. 3. System capacity vs. number of users.



References

- [DEED 78] Deed, J.E. and L.H. Gallow, "Packet Switching in a Processing Satellite", *Proc. IEEE*, Vol. 66, pp. 100-102, Jan. 1978.
- [GAVE 79] Gave, R.E., "MOMA/RM Systems with Multiple Downlink Capacity", *IEEE Trans. Com.*, vol Com-27, pp. 537-541, March 79.
- [GADR 80] Gade J. and T.E. Stern, "A Comparison of Multiple Access Protocols for Packet Switching in Satellite Switched Multi-Beam Systems", *ICC Conf. proc.*, pp 53-41-6, June 1980.
- [HHR 80] Hreniden, A. and O. Mwafi, "Analysis of a Hybrid Access Scheme for Buffered Users-Probabilistic Time Division", *MIC 1980*, pp. 71-6.1-71-6.8.
- [KLEIN 78] Kleinrock, L. and Y. Yehani, "An Optimal Adaptive Scheme for Multiple Access Broadcast Communications", *ICC June 78*, pp. 7.2.1-7.2.5.
- [LUM 77] Lum, S. S. "Delay Analysis of a Time Division Multiple Access Channel", *IEEE Trans.* Dec. 77.
- [SCHW 77] Schwartz, M., "Computer-Communication Network Design and Analysis" Prentice Hall Inc.
- [SUDA 80] Suda, T.M. Miyahara and T. Hasegawa, "Optical Bandwidth Assignment on up- and downlinks of Satellite with Buffer Capacity", *IEEE Trans. Com.* Vol. COM-28, pp. 1809-1816, Oct. 1980.





A RESERVATIONS SCHEME OF MULTIPLE ACCESS
FOR LOCAL NETWORKS

by

T. N. Saadawi
Electrical Engineering Dept.
CCNY
New York, NY

A. Ephremides
Electrical Engineering Dept.
University of Maryland
College Park, MD 20742

Abstract

A reservations scheme for multiple access suited for local networks of broadcast channels is considered and analyzed. The analysis is based on two imbedded Markov chain models, one for the individual user and one for the entire set of users. The average queue size as well as the average delay time are obtained. The proposed scheme adapts automatically to varying traffic conditions.

1. Introduction

The problem of efficient transmission of voice and computer data through shared channels remains a challenging and, in many ways, open problem in the communications area. In particular the problem of optimum or "good" design of multiple access protocols for radio channels continues to receive wide attention. There is a special need for protocols that can be analyzed so that their performance can be predicted and evaluated before implementation.

As is widely known by now, multiple access protocols may be partitioned into three main categories. The first is the class of fixed allocation protocols, like frequency division multiple access (FDMA) or time division multiple access (TDMA). These are known to be effective only for heavy, non-bursty traffic. Otherwise they are known to be extremely wasteful of channel capacity since dedicated allocation of resources is ill-suited to low duty-cycle users. The second class of protocols consists of contention-based random access schemes, such as the ALOHA protocol and its many variants. Their main disadvantages are known to be an inherent instability and a low channel utilization rate. The third broad category of protocols includes dynamic assignment schemes, and, in particular, schemes based on reservations. Few of the proposed reservation schemes have been analytically studied. Although comprehensive reviews of multiple access protocols abound in the literature

[1-3] we briefly recall some specific protocols selectively to create the context for our study. In [4] a protocol is described where the channel is divided into two subchannels; one operated in slotted ALOHA mode for reservation requests, and the other operated in a dedicated mode for data packets. Reservation schemes such as this one are particularly attractive when a significant part of channel traffic consists of multi-packet messages, because the successful transmission of a single reservation mini-packet is sufficient to reserve as many slots as are necessary for the entire message.

In [5] Crowther et al. introduced the Reservation-ALOHA idea. In this scheme, time is slotted and the slots are grouped together in fixed length frames. Slots are periodically assigned to users as in TDMA, but unlike TDMA these are temporary assignments. When any particular slot is idle in one frame, the slot is opened up for contention transmission in the following frame. A user who successfully transmits in such a contention slot will then be granted ownership of the slot as long as he has packets to transmit. This scheme was later analyzed by Lam [6].

In [7] the Round Robin reservation scheme was introduced in which a slotted channel with fixed length frame structure is assumed. Each user is permanently assigned one slot per frame, and unowned slots (which are present when the number of users is smaller than the number of slots per frame) are made available as needed to users requesting them. Users are

permitted to transmit in slots owned by terminals which currently have no packets to transmit; such inactive terminals can easily regain ownership of their slots when needed (following a one frame delay).

In [8] a reservation scheme is proposed in which channel time is divided again into frames, each of which consists of reservation time slots, preassigned (PA) time slots and reservation access (RA) time slots. In the reservation time slots each user sends a reservation message (consisting of the number of packets in his buffer minus one). Each user is assigned one PA slot permanently and a number of RA slots matching the amount requested in the reservation message.

Wieselthier and Ephremides [2] introduced the Interleaved Frame Flush-Out (IFFO) protocols, in which a long propagation delay is assumed (satellite channels). Slots are reserved by the different users, but, because of the delay, there may be slots that will go unused unless they are opened to contention. The performance of the protocols was evaluated by computation and simulation.

In this paper, we consider and analyze a simple reservation scheme essentially identical to that proposed by Rothausser and Wild [16] and studied by Bux [17] and similar to the ones proposed by Spaniol [18] and Mark [19]. This protocol has high adaptivity properties and is eminently suited to low propagation delay environments, namely, local network applications. The usual assumptions are made about user statistics, channel properties, etc. as will be described in the sequel. Our analysis is new and is based on a modeling idea introduced in [9], that consists of two coupled Markov chains, one that describes the status of the buffer contents

of a typical user (the User Markov Chain) and one that describes the status of all the users of the channel (the System Markov Chain). Such an approach permits us to account for the interaction between the users, quite similarly to the case studied in [9].

In section II, we describe the channel frame structure and the user model. In section III we describe the system Markov chain, and in section IV the user Markov chain. Delay is calculated in section V. Finally, numerical solutions and simulation results are presented in section VI. Two Appendices provide the analysis details.

II. The Model

In this section, a description of the reservation scheme is presented. This is done by first describing the channel time frame structure and then the user model.

a) The Frame structure

The frame structure is shown in Fig. (1.a), where the first slot in the frame is the reservation slot which is subdivided into M mini-slots, (M is the number of users in the system). Each user is permanently assigned one mini-slot in the reservation slot, in which he sends a request whenever he has a packet ready to transmit. In particular user i is assigned the i th mini-slot in every reservation slot. To equalize delays among the different users this assignment may be cyclically rotated so that no individual user has a permanent advantage over the others.

Mini-slots are very short since they need only accommodate the transmission of a single tone indicating the need to reserve a slot in the next frame. The length of the main frame is a random variable and depends on the number of users requesting transmission at the beginning of the frame. For example, as depicted in Fig. 1, users number 2 and number 4 are requesting transmission. Since user 1 has nothing to transmit, user 2 knows he will be the first to transmit after the reservation slot is over. Also, since user 3 did not request transmission, user 4 knows he will transmit after user 2, and then the new frame will start. Thus, the frame length equals to three slots. Note that the frame length cannot be less than one slot and cannot be more than $(M+1)$ slots (corresponding to the case that all M users are requesting transmission). We assume that messages consist of single packets each of which fits exactly in one slot. Also we assume that the channel is monitored by all users without delay or noise. Note also, that a user transmits only one packet per frame regardless of the number of packets in his buffer.

b) The User Model

The system consists of M terminals, each of which has an infinite size buffer. The arrival process at each of the M terminals is assumed to be a Bernoulli process with rate c . The user may be in one of two states, an idle state, if his buffer is empty or an active state, if his buffer is non-empty.

As shown in Fig. 1.b, whenever the user has a packet in his buffer, he sends a reservation request at the beginning of

the frame, and at the same time the packet at the head of line (HOL) is moved to the transmitter and waits there until it is transmitted in its assigned slot.

We need the following definitions:

Let

$\pi_i \triangleq$ steady state probability of having i packets in the buffer at the beginning of a frame.

Hence

$\pi_0 =$ probability of a buffer being empty
 $=$ probability of a user being idle.

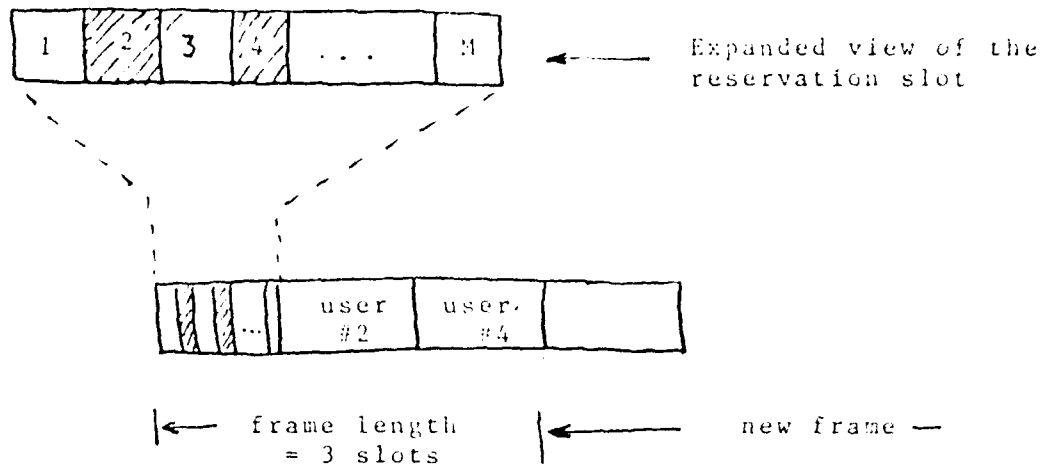
$1 - \pi_0 =$ probability of a user being active.

Such a model produces a non-standard queueing system of interacting queues. The service time of a packet in one queue depends on the status of the other queues. If we let n_k represent the contents of the input buffer of the k^{th} user, then a complete description of all M users requires the specification of the joint probability distribution $P(n_1, n_2, \dots, n_M)$. The determination of these probabilities is very complex, if not impossible, and demands the solution of a large number of sets of equations. Fayolle and Iasnogorodski [10] showed the limitation of this direct approach. They considered the simple example of two exponential servers the service rate of each of which depended on the status of the other server's queue. Using the theory of analytic continuation they reduced the

determination of the joint probability distribution of the two queues to a Riemann-Hilbert problem and were able to obtain a very complex closed form for the solution. Furthermore extension of their analysis to the case of more than two users is not feasible. Leibowitz [11] presented an approximate method of treating multiqueue systems. He studied the case of N queues with a single server, in which the queues are served in cyclic order. He made the assumption that the distribution of the queue size of one queue is invariant if it is the same for all queues during one cycle.

Hashida [12, 13] used Leibowitz's approximation in the analysis of multiqueue systems. Also in [8], the authors used Leibowitz's approximation along with an independence assumption.

We consider that the main contribution of our paper is the proposed mathematical model for the analysis of interacting buffered terminals. The model, as we mentioned earlier, is basically a combination of two Markov chains imbedded at the beginning of each frame; one Markov chain that describes the state of the user, referred to as the user Markov chain, and another Markov chain that describes the state of all the users in the system, referred to as the system Markov chain.



(a) Typical frame structure for the reservation scheme



(b) The User Model

Fig. 1

III The System Markov Chain

The state of the system is described by the number of users requesting transmission at the beginning of the frame. First, we need the following definitions: Let

j = number of users requesting transmission at the beginning of a frame.

i_j = frame length when j users request transmission at the beginning of the frame

$$= 1 + j$$

$p_j \triangleq$ steady state probability of j users requesting transmission at the beginning of the frame

$\sigma_j \triangleq$ Probability an idle user generates at least one packet during a frame of length $(1+j)$ slots

$$= 1 - (1 - \sigma)^{j+1}$$

F = probability of one packet only in the buffer at the beginning of a frame given user is active

$$= \frac{\pi_1}{1 - \pi_0}$$

The system Markov chain is illustrated in Fig. 2.a. while its transition probabilities are obtained in appendix A.

IV The User Markov Chain

The state of the user is described by the number of packets in the buffer at the beginning of the frame. Let

n = number of packets in buffer at the beginning of a frame

π_n = steady state probability of n packets in the buffer at the beginning of a frame

Let $\pi(z)$ be the generating function of π_n , i.e.

$$\pi(z) = \sum_{n=0}^{\infty} \pi_n z^n$$

In Fig. 2.b, we show the user Markov chain. Note that only one packet can leave the buffer per frame. We show the derivation of the transition probabilities in Appendix B.

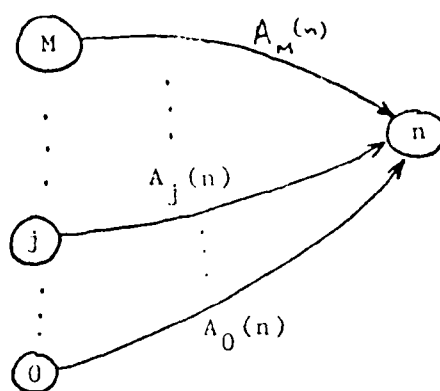


Fig. 2.a System Markov Chain

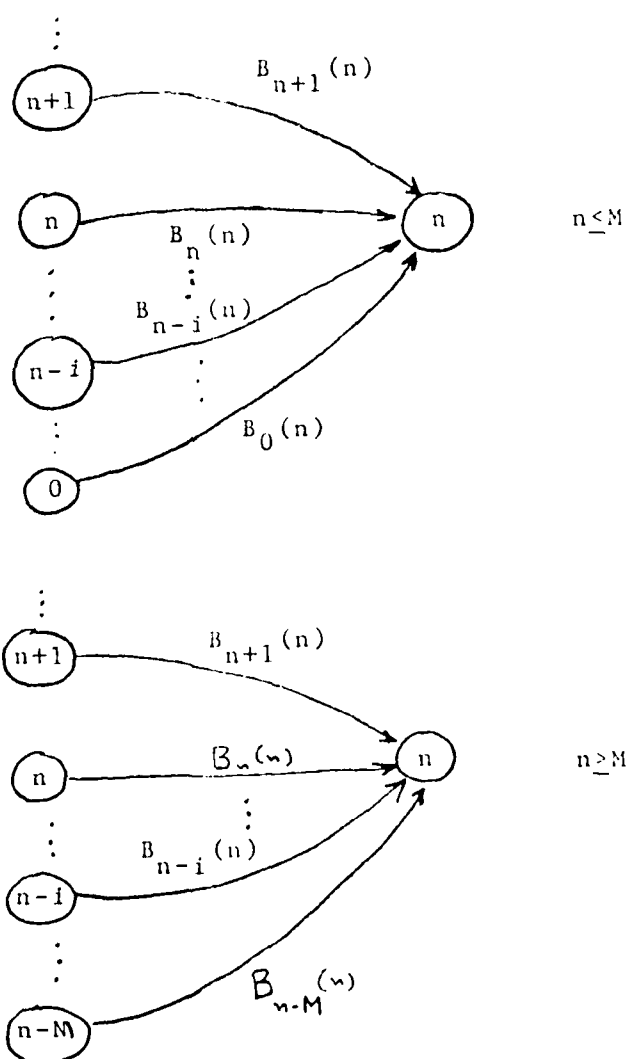


Fig. 2.b User Markov Chain

V Delay Analysis

The average total delay per packet, D , is the sum of the average waiting time in the buffer, W , plus the average time from the beginning of the frame during which the packet is scheduled for transmission until the packet is transmitted in its assigned slot. We refer to the latter as the service time, S .

Hence,

$$D = W + S$$

Using Little's result, we can write W as;

$$W = \frac{Q}{\sigma}$$

where, Q is the average queue size.

Assuming the cyclic assignment discussed earlier in section II, the average service time equals to one plus half the average number of requesting users at the beginning of the frame, i.e.

$$S = 1 + \frac{1}{2} \bar{j}$$

VI Numerical Solutions

Equations (A-1) through (A-5) provide the values of the quantities that were needed in order to solve the state transition equations for the system probabilities p_n . However, some of these quantities are not solely expressed in terms of the parameters M and σ , but also in terms of F , i.e., the probability of one packet in the buffer at the beginning of a frame given user is active, which is a function of π_1 and π_0 . But, π_0 and π_1 both depend in turn on $\{p_n\}$. Thus, in total, we have a set of simultaneous coupled non-linear equations in π_0 , π_1 and $\{p_n\}$. In Fig. 3, we show the flow chart for the numerical procedure used to solve these equations. We start with an initial value for F , then solve $(M+1)$ linear simultaneous equations in $\{p_n\}$. We then determine π_0 and π_1 , and use Eq. (B.1) and Wegestein's iteration

scheme [14] to obtain the new value of F . Two to five iterations were required for convergence to a solution for F within a 10^{-7} tolerance in the values of F and $\{p_n\}$.

For $M=4$, Fig. 4 shows the probabilities π_0 (the prob. of an empty buffer), p_0 (the prob. of zero users requesting transmission) and p_M (the prob. of M users requesting transmission) versus channel throughput. Notice that as throughput increases both π_0 and p_0 decrease while p_M increases. Figs. 5 and 6 show the average waiting, service, and total times versus throughput for $M=4$ and 8 respectively. Throughput is defined as $\lambda = M\sigma$. Such a definition presupposes steady-state stability so that the average input rate is equal to the average transmission rate. Although detailed stability analysis is not considered here, it is fairly obvious that since there are no wasted slots (due to contention or idleness) except for one reservation slot in each frame and since all users are assumed similar, any value of λ less than $\frac{M}{M+1}$ produces a stable equilibrium.

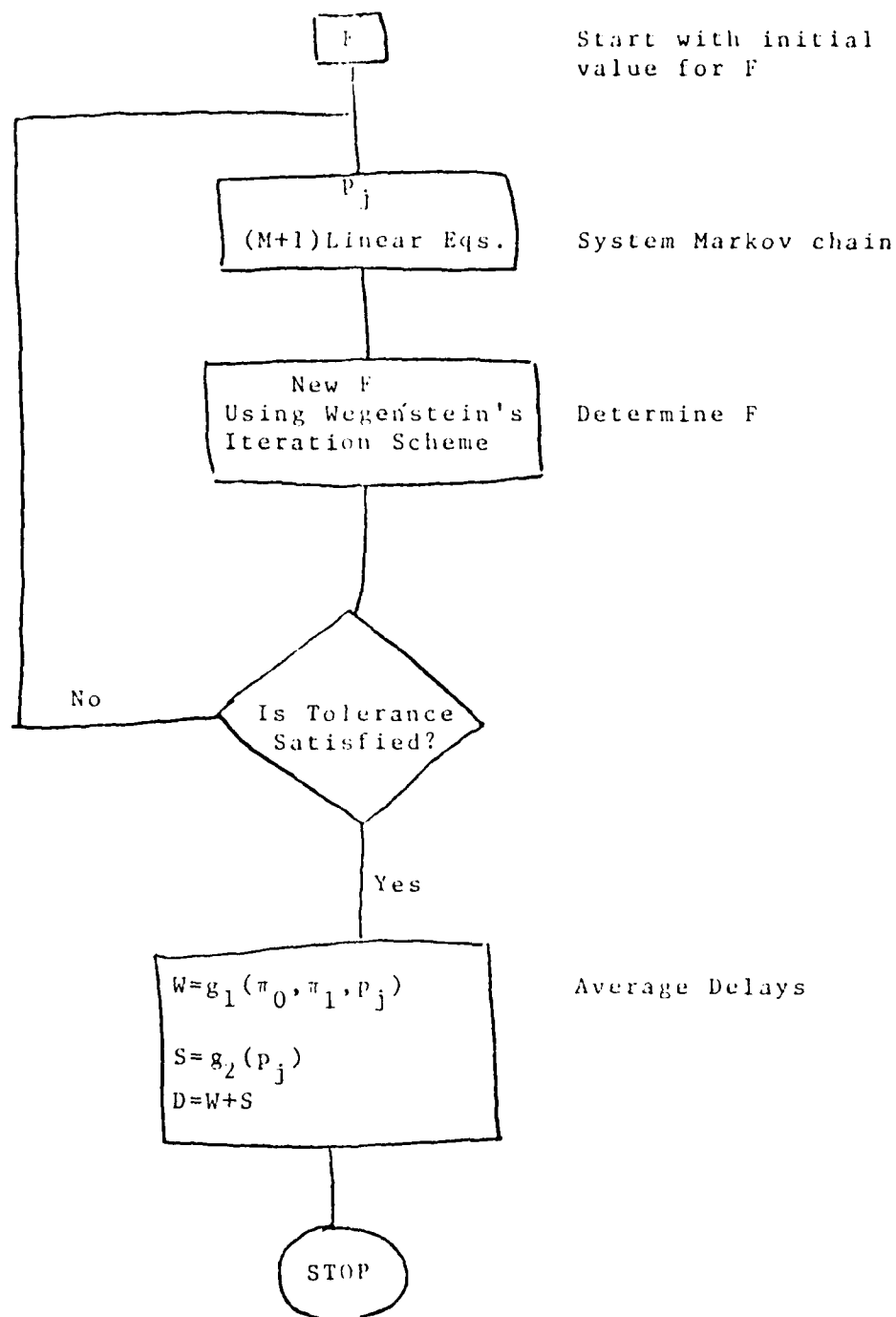


Fig. 3. Flow Chart to Determine Average Delays for the Reservation Scheme

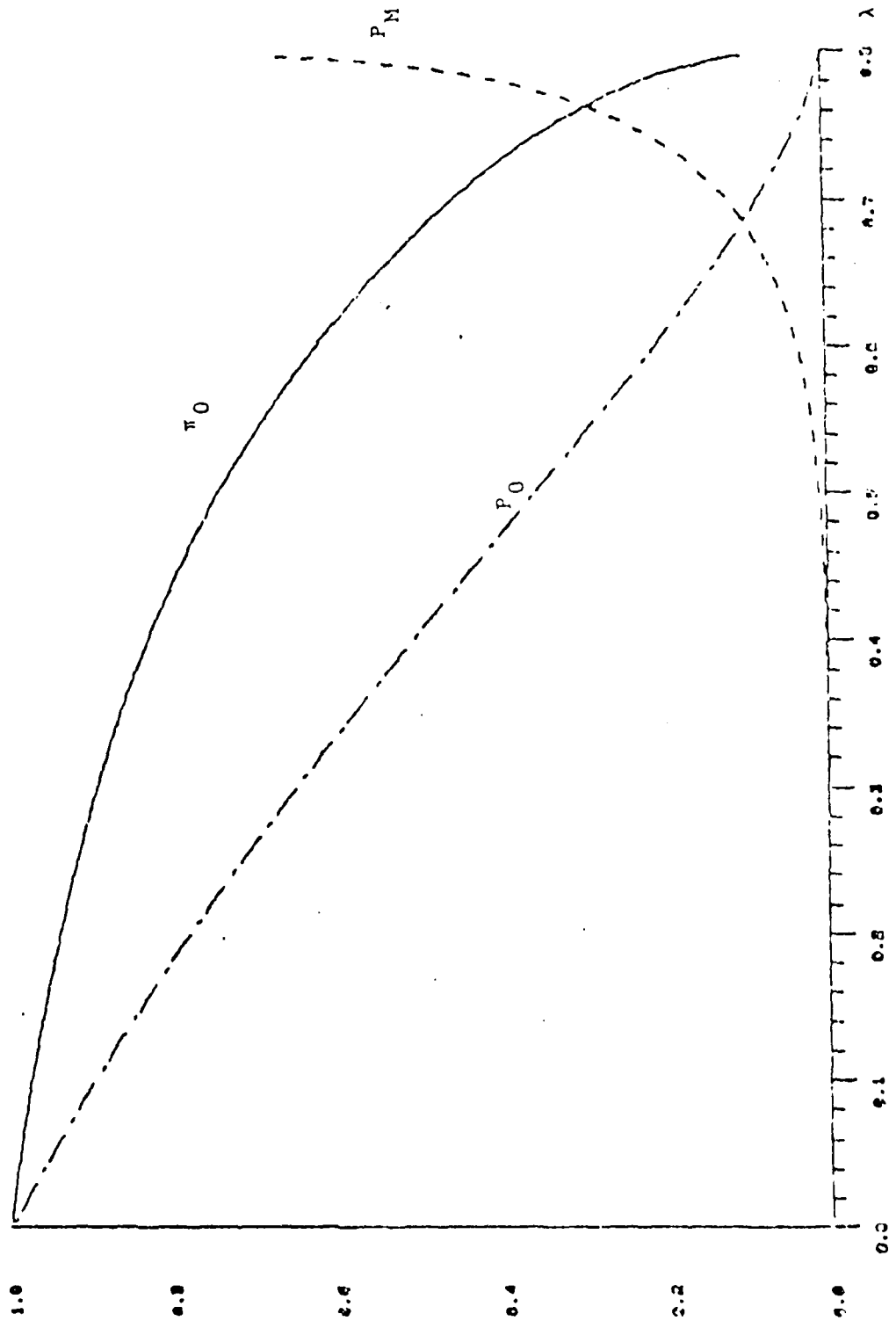


Fig. 44. Probabilities π_0, P_0, P_M versus Throughput

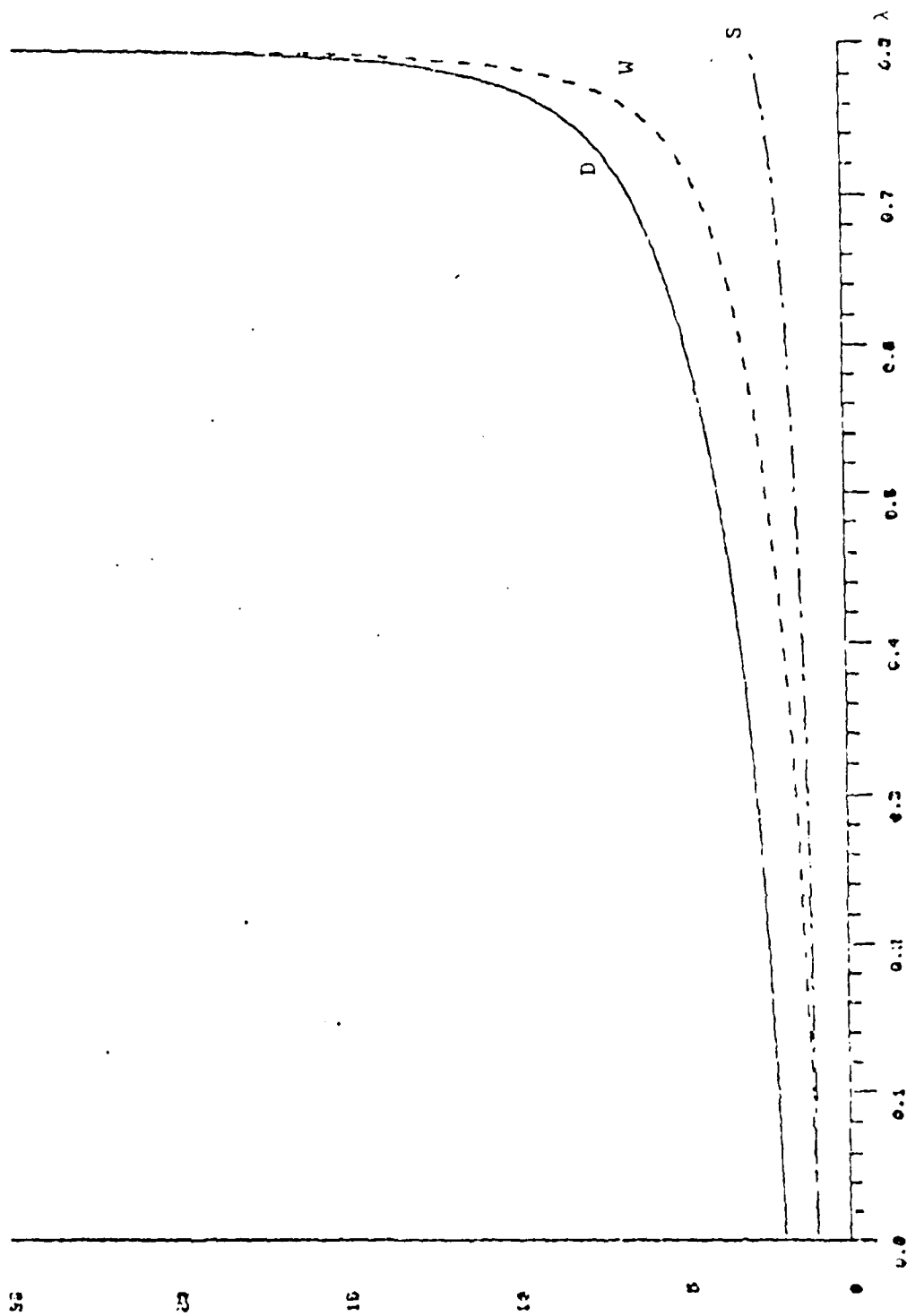


Fig. 5. Average Delays -vs- Throughput
 $M=4$

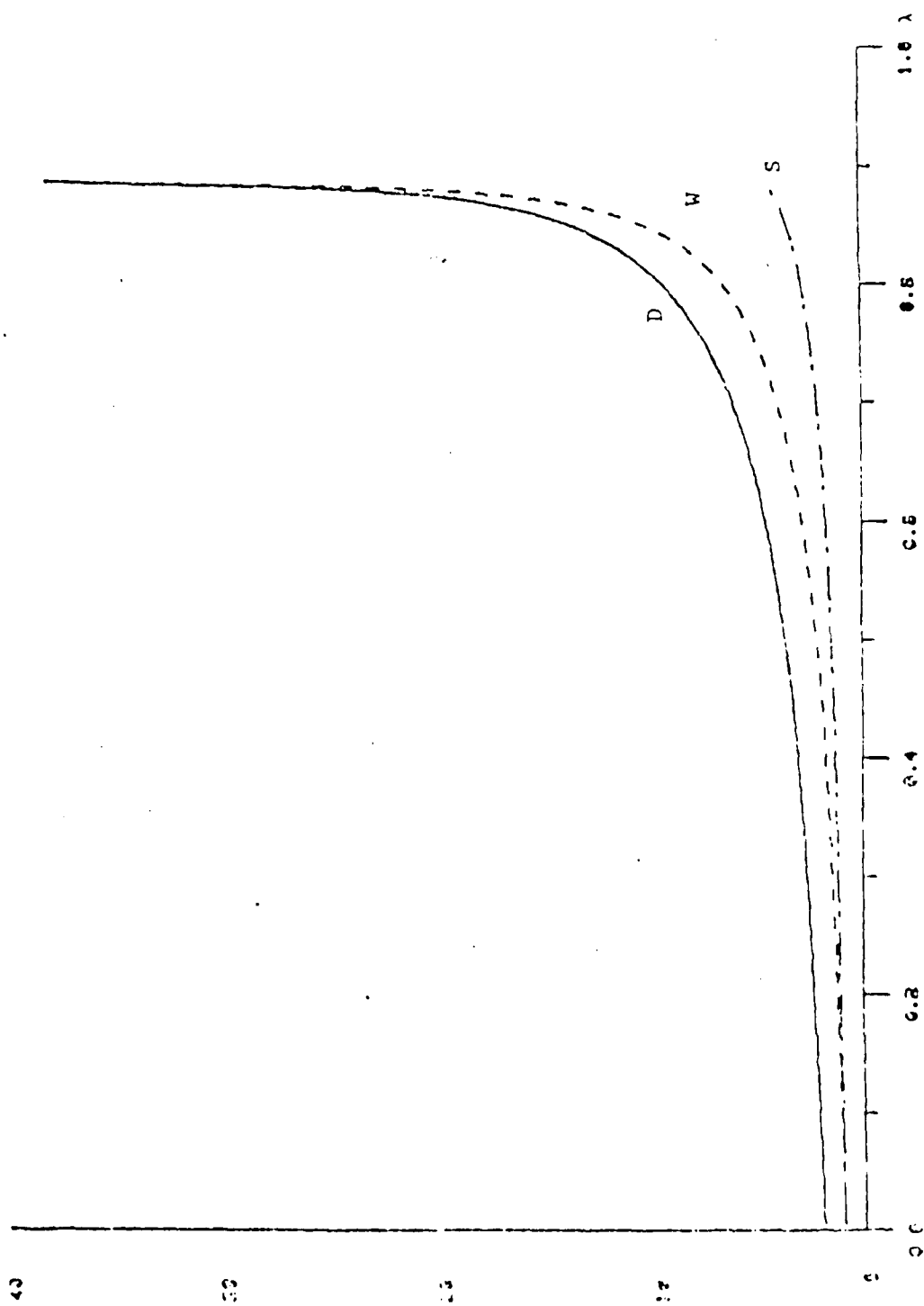


Fig. 6. Average Delays -vs- Throughput
 $M=8$

The reservation scheme has been simulated on the computer. The simulation program was written in FORTRAN, it was not necessary to use a language specifically designed for simulation. The numerical solution for the reservation scheme required the implicit assumption that some processes are independent when in fact they are not. The simulation produced agreement with the numerical results within 5% for the average queue size. Table 1 shows the comparison between numerical and simulated results for average waiting, service and total delays for four users. The maximum difference in the average waiting time is approximately 6% and in the average service time is 5%.

Throughput		D	W	S
.10199	A	2.2016	1.1448	1.0568
	S	2.1687	1.1414	1.0560
.30597	A	2.812	1.5905	1.2215
	S	2.7613	1.5419	1.2194
.50995	A	3.9797	2.4602	1.5195
	S	3.8001	2.3691	1.5194
.71393	A	6.8565	4.7208	2.1357
	S	6.7625	4.5175	2.245

A: Analytical results

S: Simulated results

Table 1. Comparison of Numerical and Stimulated Results for the Average Delays $M = 4$

VI Conclusion

In this paper, we have introduced and analyzed a reservation scheme, in which reservation packets are broadcast in TDMA fashion over a fraction of the main channel. The average waiting, service and total times have been obtained. The analysis is based on a combination of two imbedded Markov chains, that partially account for the interdependence between the users. Such an approach has been implemented by the authors in the analysis of slotted-ALOHA with a finite number of buffered users [9]. Also in [13], we generalized the analysis of the reservation scheme presented in this paper for a class of Collision Resolution Algorithms of the Capetanakis type assuming a finite number of buffered users.

APPENDIX A

System Markov Chain for the Reservation Scheme

Let

K = number of active users out of the j requesting users

The distinction between the number of active users and the number of users requesting transmission is necessitated by the assumption that if there is only one packet in the buffer, it is automatically moved to the transmitter box at the beginning of the frame and hence the corresponding user is considered idle at the beginning of that frame. Thus

$M-K$ = number of idle users at the beginning of a frame

Recall that F is defined as;

F = Probability of one packet only in buffer at the beginning of a frame given user is active

$$= \frac{\pi_1}{1-\pi_0} \quad (A.1)$$

Let

$C_K(j)$ = Prob. [K active users out of j requesting users at the beginning of a frame]

$$= \binom{j}{K} (1-F)^K (F)^{j-K} \quad (A.2)$$

In order to use the above binomial distribution, it is assumed that the event that a user is active is independent from the state of the other users. Similar assumption has been stated and justified in [13] for the case of a single server serving N lines in cyclic order.

Let

$q(i|K, j) \triangleq \text{Pr}[i \text{ of the } M-K \text{ idle users generate packets in a frame given that } j \text{ users are requesting and } K \text{ of these } j \text{ are active}]$

$$= \binom{M-K}{i} c_j^i \{1-c_j\}^{M-K-i} \quad (A.3)$$

Now we can write $h(i, K|j)$, the joint probability distribution of i and K conditioned with respect to j , as follows:

$h(i, K|j) \triangleq \Pr\{i \text{ out of } M-K \text{ idle users generate packets during a frame and } K \text{ out of the } j \text{ requesting users are active given } j \text{ users are requesting}\}$

$$\begin{aligned} &= \binom{M-K}{i} \sigma_j^i [1-\sigma_j]^{M-K-i} \binom{j}{K} (1-F)^K (F)^{j-K} \\ &= q(i, K, j) C_K(j) \end{aligned} \quad (A.4)$$

Referring to Fig. (2.a) we can write $A_j(n)$, the transition probability from state j to state n , as;

$$A_j(n) = \sum_{K=0}^{\min(j, n)} \Pr\{(n-K) \text{ arrivals and } K \text{ users active given } j \text{ users are requesting}\}$$

$$= \begin{cases} \sum_{K=0}^{\min(j, n)} h(n-K, K, j), & 1 \leq j \leq M, \quad 0 \leq n \leq M \\ \binom{M}{n} \sigma^n (1-\sigma)^{M-n}, & j = 0, \quad 0 \leq n \leq M \end{cases} \quad (A.5)$$

Of course then we have

$$p_n = \sum_{j=0}^M A_j(n) p_j, \quad 0 \leq n \leq M$$

APPENDIX B

The User Markov Chain for the Reservation Scheme

In this Appendix, we determine the transition probabilities for the user Markov chain shown in Fig. 2.b. Also, expressions for p_0 , the probability of an empty buffer and Q , the buffer size are obtained. First, we need the following definitions; Let

$g(i, j+1) \triangleq$ Probability that i out of $j+1$ possible packets arrive during a frame given j requesting users at the beginning of the frame

$$= \binom{j+1}{i} c^i (1-c)^{j+1-i}$$

$$H(z, j+1) \triangleq \sum_{i=0}^{j+1} g(i, j+1) z^i = [1-c(1-z)]^{j+1} \quad (B.1)$$

$$H(z) \triangleq \sum_{j=0}^M H(z, j+1) p_j \quad (B.2)$$

Now, referring to Fig. 2.b, we have;

$$\begin{aligned} B_{n+1}(n) &= \sum_j \text{Prob. [one packet leaves and no packets arrive} \\ &\quad \text{given } j \text{ requesting users at the beginning} \\ &\quad \text{of a frame]} \\ &= \left[\sum_{j=1}^M g(0, j+1) p_j \right] / (1-p_0) \quad n = 0, 1, \dots \end{aligned} \quad (B.3)$$

Notice that the initial value of the summation is $j = 1$, since at the beginning of the frame we have at least one user requesting transmission. For the same reason, we normalize by dividing by $(1-p_0)$.

Similarly,

$$\begin{aligned} B_n(n) &= \sum_j \text{Pr. [one packet leaves, and one arrives during a} \\ &\quad \text{frame given } j \text{ users requesting at the beginning} \\ &\quad \text{of the frame]} \end{aligned}$$

$$= \begin{cases} \sum_{j=1}^M g(1, j+1) p_j / (1-p_0) & , \quad n \neq 0 \\ \sum_{j=0}^{M-1} g(0, j+1) p_j / (1-p_M) & , \quad n=0 \end{cases} \quad (B.4)$$

Note that for $n=0$, we can have a maximum of $(M-1)$ requesting users; hence, we need to normalize by dividing by the probability of having at most $(M-1)$ requesting users.

Similarly,

$$\begin{aligned} B_{n-i}(n) &= \sum_j \text{Pr. [one packet leaves, } i+1 \text{ arrive given } j \\ &\quad \text{requesting users]} \\ &= \sum_{j=i}^M g(i+1, j+1) p_j / (1-p_0) \quad n = 2, 3, \dots \\ &\quad i = 1, \min(n-1, M) \end{aligned} \quad (B.5)$$

and

$$B_0(n) = \sum_{j=n-1}^{M-1} g(n, j+1) p_j / (1-p_M) \quad n = 1, \dots, M \quad (B.6)$$

Hence,

$$\pi_n = \begin{cases} B_{n+1}(n) \pi_{n+1} + B_n(n) \pi_n + \sum_{i=1}^{n-1} B_{n-i}(n) \pi_{n-i} + B_0(n) \pi_0, & 2 \leq n \leq M \\ B_{n+1}(n) \pi_{n+1} + B_n(n) \pi_n + \sum_{i=1}^M B_{n-i}(n) \pi_{n-i}, & n > M \end{cases} \quad (B.7a)$$

$$\pi_0 = B_1(0) \pi_1 + B_0(0) \pi_0 \quad (B.7b)$$

$$\pi_1 = B_2(1) \pi_2 + B_1(1) \pi_1 + B_0(1) \pi_0 \quad (B.7c)$$

To obtain $\pi(z)$, we need to define the following; let

$$G(0) \triangleq \sum_{j=0}^M g(0, j+1) p_j \quad (B.8)$$

$$G(1) \triangleq \sum_{j=0}^M g(1, j+1) p_j \quad (B.9)$$

Multiplying Eq. (B.7) by z^n and summing over n , we get;

$$\begin{aligned}\pi(z) &= \sum_{n=0}^{\infty} \pi_n z^n \\ &= \sum_{n=0}^{\infty} B_{n+1}(n) \pi_{n+1} z^n + \sum_{n=0}^{\infty} B_n(n) \pi_n z^n + \sum_{n=0}^{\infty} \sum_{i=1}^{\min(n-1, M)} B_{n-i}(n) \pi_{n-i} z^n u(n-i) \\ &\quad + \sum_{n=0}^{\infty} B_0(n) \pi_0 z^n u(n-1)\end{aligned}\quad (B.10)$$

Using Eq. (B.3), we get

$$\sum_{n=0}^{\infty} B_{n+1}(n) \pi_{n+1} z^n = \frac{1}{1-p_0} [G(0) - g(0,1)p_0] z^{-1} [\pi(z) - \pi_0] \quad (B.11)$$

Using Eq. (B.4), we get;

$$\sum_{n=0}^{\infty} B_n(n) \pi_n z^n = \frac{1}{(1-p_0)} [G(1) - g(1,1)p_0] [\pi(z) - \pi_0] + \frac{[G(0) - g(0, M+1)p_M]}{(1-p_M)} \pi_0 \quad (B.12)$$

From (B.6) we get;

$$\begin{aligned}\sum_{n=1}^M B_0(n) \pi_0 z^n &= \sum_{j=1}^M \sum_{n=1}^j g(n, j+1) z^n p_j \pi_0 / (1-p_M) \\ &= \frac{\pi_0}{(1-p_M)} [H(z) - G(0) - H(z, M+1)p_M + g(0, M+1)p_M]\end{aligned}\quad (B.13)$$

Finally, we have

$$\begin{aligned}\sum_{n=2}^{\infty} \sum_{i=1}^{\min(n-1, M)} B_{n-i}(n) \pi_{n-i} z^n &= \sum_{i=1}^M \sum_{n=i+1}^{\infty} B_{n-i}(n) \pi_{n-i} z^n \\ &= \frac{1}{1-p_0} \sum_{i=1}^M \sum_{j=i}^M g(i+1, j+1) p_j z^i [\pi(z) - \pi_0] \\ &= \frac{1}{1-p_0} \sum_{j=1}^M \sum_{i=1}^j g(i+1, j+1) p_j z^i [\pi(z) - \pi_0]\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{1-p_0} \sum_{j=1}^M z^{-1} [H(z, j+1) p_j - g(1, j+1) p_j z - g(0, j+1) p_j] [\pi(z) - \pi_0] \\
&= \frac{z^{-1}}{(1-p_0)} [\pi(z) - \pi_0] [H(z) - zG(1) - G(0)]
\end{aligned} \tag{B.14}$$

From (B.11) - (B.14) and (B.10) we obtain

$$\pi(z) = \frac{X(z)}{Y(z)} \pi_0 \tag{B.15}$$

where,

$$X(z) = \left[\frac{1-p_0}{1-p_M} z - 1 \right] H(z) + p_0 (g(0,1) + zg(1,1)) - \frac{(1-p_0)}{(1-p_M)} H(z, M+1) p_M \tag{B.16}$$

$$Y(z) = (1-p_0)z - H(z) + p_0 (g(0,1) + zg(1,1)) \tag{B.17}$$

To obtain π_0 , we take the limit in Eq. (B.15) as $z \rightarrow 1$, hence,

$$\pi_0 = \frac{\dot{Y}(1)}{\dot{X}(1)} \tag{B.18}$$

where

$$\dot{Y}(1) = \left. \frac{\partial Y(z)}{\partial z} \right|_{z=1} = 1 - p_0(1-\sigma) - \dot{H}(1) \tag{B.19}$$

$$\dot{X}(1) = \frac{1}{(1-p_M)} \{ (p_M - p_0) \dot{H}(1) + 1 - p_0 + \sigma p_0 (1-p_M) - (1-p_0) p_M ((M+1)\sigma + 1) \} \tag{B.20}$$

To obtain Q , we differentiate (B.15) and take the limit as $z \rightarrow 1$;

Thus

$$Q = \frac{\ddot{X}(1)\pi_0 - \ddot{Y}(1)}{2\dot{Y}(1)} \tag{B.21}$$

where

$$\begin{aligned}
\ddot{X}(1) = \left. \frac{\partial^2 X(z)}{\partial z^2} \right|_{z=1} &= \frac{1}{(1-p_M)} \{ (p_M - p_0) \ddot{H}(1) + 2(1-p_0) \dot{H}(1) - (1-p_0) p_M \\
&\quad [\ddot{H}(1, M+1) + 2\dot{H}(1, M+1)] \}
\end{aligned} \tag{B.22}$$

and

$$\ddot{Y}(1) = \ddot{H}(1) \tag{B.23}$$

References

- [1] Tobagi, F., "Multiaccess Protocols in Packet Communication Systems", IEEE Trans. on Communications, Vol. 28, pp. 468-488, April 1980.
- [2] Wieselthier, J.E. and Ephremides, A., "A New Class of Protocols for Multiple Access in Satellite Networks", IEEE Transactions on Automatic Control, October 1980.
- [3] Kleinrock, L., Queueing Systems, John Wiley, 1976.
- [4] Roberts, L.G., "Dynamic Allocation of Satellite Capacity Through Packet Reservation," AFIPS Conference Proceedings, 1973 National Computer Conference, 42, 711-716.
- [5] Crowther, W., R. Rettberg, D. Walden, S. Ornstein, and F. Heart, "A System for Broadcast Communications Reservation-ALOHA," Proceedings of the Sixth Hawaii International Conference on System Sciences, University of Hawaii, Honolulu, January 1973.
- [6] Lam, S.S., "Packet Broadcast Networks - A Performance Analysis of the R-ALOHA Protocol", IEEE Transactions on Computers, Vol. C-29, No. 7, July 1980.
- [7] Binder, R., "A dynamic packet switching system for Satellite broadcasting channels" Proceedings of the IEEE International Conference on Communications, Vol. 3, June 1975.
- [8] Balagangadhar and R.L. Pickholtz, "Analysis of a Reservation Multiple Access Technique for Data Transmission Via Satellites", IEEE Trans. on Communications, vol. COM-27, No. 10, October 1979.
- [9] Saadawi, T.N. and A. Ephremides, "Analysis, stability and Optimization of Slotted ALOHA with a Finite Number of Buffered Users", IEEE Trans. on AC, June 1981.

- [10] Fayolle, G. and Iasnogorodski, "Two Coupled Processors: The Reduction to a Riemann-Hilbert Problem" Wahrscheinlichkeits Theorie, Springer-Verlag, 1979.
- [11] Leibowitz, M.A., "An Approximate Method for Treating a Class of Multiqueue Problems," IBM Journal of Research and Development, Vol. 5, No. 3, 204-209, July 1981.
- [12] Hashida, O. and K. Ohara, "Line Accomodation Capacity of a Communication Control Unit" Review of the Electrical Communication Laboratories, Vol. 20, No. 3-4, March-April, 1972.
- [13] Hashida, O., "Analysis of Multiqueue", Review of the Electrical Comm. Lab., Nippon Telegraph and Telephone Public Corporation, Vol. 20, No. 3-4, March-April, 1972.
- [14] Lance, G.N., Numerical Methods for High Speed Computers, Iliffe, London, 1960, pp. 134-138.
- [15] Saadawi, T.N. and A. Ephremides, "Analysis of the Tree Algorithm with a Finite Number of Buffered Users", Proceedings ICC, June 1981.
- [16] Rothaus, E., and Wild, D., "MLMA: A Collision-free Multiaccess Method", Proc. IFIP Congress 77, Amsterdam, 1977, pp. 431-436.
- [17] Bux, W., "Local-Area Subnetworks = A Performance Comparison", IEEE Trans. on Comm., Vol. 29, No. 10, October 1981, pp. 1465-1473
- [18] Spaniol, O., "Modeling of Local Computer Networks", Computer Networks, Vol. 3, pp. 315-326, 1979.
- [19] Mark, J.W., "Global Scheduling Approach to Conflict-Free Multiaccess via a Data Bus", IEEE Trans. on Comm., Vol. COM-26, No. 9, Sept. '78, pp. 1342-1352.

AN ALGORITHM FOR SCENE MATCHING

By

Tarek N. Saadawi and George Eichmann

Abstract

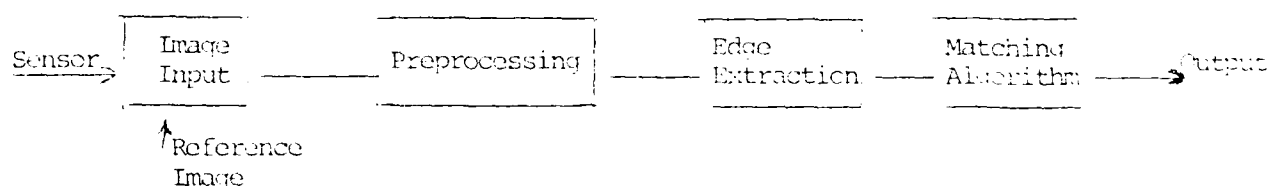
A fast algorithm for scene matching is studied and compared with other algorithms. The measure of similarity is the sum of the absolute values of the differences of image intensity between the template and the scene. Comparison with other measures of similarity is presented.

(I) Introduction

The need for analyzing the images of the earth taken by various sensors on high and low altitude aircraft is increasing at a rapid pace. These images have been used for several purposes, including navigational guidance, geographic and topographic map matching, natural resource analysis, weather prediction, and environment studies.

A common problem that arises in these applications is scene matching. Given a pictorial description of a region of scene, determine which region in another image is similar. The method used to solve this problem in its simplest form is called template matching. An image of template is given, and it is desired to determine all locations of the template in another image. If a sufficient number of corresponding regions could be determined then the image could be accurately registered, facilitating location of corresponding objects.

Including the basic template matching method, several approaches and techniques have been used, such as sequential similarity comparisons, matching with invariant moments, and matching with edge features. An image edge, as defined here, is a local change or discontinuity in image luminance. Generally, image matching can be shown in the following functional diagram.



Preprocessing is an important procedure in image matching. Its basic objective is correcting two degraded images (such as two images from two different sensors) to get new images to meet certain requirements of accuracy and projection. Generally, it includes geometrical correction and intensity correction. Geometrical correction is a transform to change each pixel of distorted image from original coordinates (x, y) to geometrically more corrected coordinates (u, v) . For intensity correction, a matching transform-

ation method called Karhonen-Loeve transform was widely used.

Edge information is important for both human and machine perception and a considerable amount of research has been done in attempts to extract edge information from pictures.

In this paper we study an algorithm for image matching. The algorithm requires less computation time and achieves a high registration performance. In the next section we discuss the algorithm and in Section III we analyze the results. Section IV compares the performances of the different measures of similarity for image registration. The conclusion is given in Section V.

11 An Algorithm for Scene Matching

Sullivan and Martin, in (1), proposed a coarse search correlation technique for image registration, the sequential similarity detection algorithm (SSDA). They used the normalized correlation coefficient as the measure of match. The normalized correlation coefficient is defined as

$$C(i,j) = \frac{M^2 c_1 - 4 c_4 c_5}{\{(M^2 c_1 - c_4^2)(M^2 c_3 - c_5^2)\}^{1/2}} \quad 0 \leq i, j \leq L - M + 1 \quad (1-a)$$

where

$$\begin{aligned} c_1 &= \sum_{p=1}^M \sum_{q=1}^M \{S^{ij}(p,q)\}^2 \\ c_2 &= \sum_{p=1}^M \sum_{q=1}^M S^{ij}(p,q) \cdot T(p,q) \\ c_3 &= \sum_{p=1}^M \sum_{q=1}^M \{T(p,q)\}^2 \\ c_4 &= \sum_{p=1}^M \sum_{q=1}^M S^{ij}(p,q) \\ c_5 &= \sum_{p=1}^M \sum_{q=1}^M T(p,q) \end{aligned} \quad (1-b)$$

Where $T(p,q)$ denotes the template with dimension $M \times M$ pixels, $S^{ij}(p,q)$ is $M \times M$ subimage of the scene located at (i,j) coordinates with dimension $L \times L$. Hence the search area would be $(L - M + 1) \times (L - M + 1)$. The formula in equation (1) requires a large computation time.

A critical issue is how to choose a thresholding technique for terminating the calculation of $C(i,j)$ early. If the thresholds are too high, we may eliminate the best registration because of the early behavior of the calculation of $C(i,j)$. On the other hand, if the thresholds are too low, many poor registrations will avoid early termination of eqn. (1). The authors in (1) proposed an empirically derived function as the threshold sequence dependent on the subset number b .

Here, we investigate using the sum of the absolute values of the differences between two images as the measure of similarity. The image similarity measurement $A(i,j)$ is defined as

$$A(i,j) = \frac{1}{M} \sum_{p=1}^M \sum_{q=1}^M |S^{ij}(p,q) - T(p,q)| \quad (2)$$

The calculation procedures are similar with the SSDA algorithm. First, we use a subset of the two data fields (S^{ij} and T) at a low resolution to calculate an estimate of $A_p(i,j)$. For example, $A_1(i,j)$ would be calculated using one-sixteenth of the data. This $A_1(i,j)$ could then be compared to a threshold. If it passes the threshold, the next estimate $A_2(i,j)$ at the next high level resolution is calculated using an additional one-sixteenth of the data. This procedure is continued until all of the data points are used, giving the best estimate $A_p(i,j)$ or until a threshold is not exceeded, which aborts the evaluation process.

The sampling rule of the data set is shown in Fig. 1. The coarse spacing is taken in a pseudo-random manner. Each pass requires that one additional pixel be chosen from each block. The choice of the sampling method is important and affects the matching procedure. This will be studied in the following section.

The threshold sequence is taken as a constant value. It is shown that the threshold constant is between 30% to 50% of the average value of the intensity per pixel in the template image.

1.1 Results

In this section a simple example will be presented to demonstrate the new algorithm implementation.

A) The Template and the Scene

Computer generated images (English letters) are used in this example. Also, the noise is generated by the computer (Gaussian noise). The template and the scene have the same object (letter "F"), but they have different intensity distribution and background noise.

The template consists of character "F" and its dimension is 16x16. Its intensity distribution has been quantized into 128 levels. Within "F", the intensity has Gaussian distribution plus Gaussian noise with zero mean and $\sigma = 43$, the background (outside "F") has random noise with intensity quantized levels between 0 and 30.

Within "F" the intensity distribution can be expressed as

$$I_{Tp}(x,y) = 128 \times \sum_{k=1}^3 \left\{ \exp\left[-(x-x_k)^2 / \Delta x_k^2\right] \cdot \exp\left[-(y-y_k)^2 / \Delta y_k^2\right] \cdot \text{rect}\left[(x-x_k)/\Delta x_k\right] \cdot \text{rect}\left[(y-y_k)/\Delta y_k\right] \right\} \quad (3a)$$

where "F" is broken into K blocks ($K = 3$)

$\text{rect}(X/\Delta x)$ function is defined as the "slit" function;

$$\text{rect}(X/\Delta x) = \begin{cases} 1 & |x| \leq \Delta x/2 \\ 0 & |x| > \frac{\Delta x}{2} \end{cases}$$

and the Gaussian noise has the form

$$f_{ng}(I_{Tp}) = \text{EXP}\left[-I_{Tp}^2 / (\sigma^2 \times 43 \times 43)\right] \quad (3b)$$

Outside "F," the background noise distribution is

$$I_{nb} = \text{INT}\left\{ 150 \times \left[1 + 2 \times \left| \text{RAN}(1) - 0.5 \right| \right] \right\} \quad (3c)$$

Now, we express the template as

$$I_T = I_{Tp} \left[1 + 0.5 \times f_{ng} \right] + I_{nb} \quad (4)$$

four scene images have been generated (letters: "I," "L," "E," and "X").

The scene image (I_{sc}) can be expressed as

$$I_{sc} = 128 \times \sum_{k=1}^9 \exp \left[-\frac{(x-x_k)^2}{2\sigma_k^2} \right] \exp \left[-\frac{(y-y_k)^2}{2\sigma_k^2} \right] \cdot \text{rect} \left[\frac{x-x_k}{\Delta x_k} \right] \cdot \text{rect} \left[\frac{y-y_k}{\Delta y_k} \right] \\ + 128 \times \sum_{k=1}^4 \exp \left[-\frac{(x-x_k)^2}{2\sigma_k^2} \right] \exp \left[-\frac{(y-y_k)^2}{2\sigma_k^2} \right] \text{rect} \left[\frac{x-x_k}{\Delta x_k} \right] \text{rect} \left[\frac{y-y_k}{\Delta y_k} \right]$$

where "P," "E," "L" are broken into 3, 4, and 2 blocks respectively; and "X" into 4 blocks.

Considering the noise, then

$$I_s = I_{sc}(1 + 0.5 \times f_{ng}) + I_{nb} \quad (5)$$

The template and scene images are shown in figures 2a and 2b. The total and average intensity of both the template and scene are given in Table 1. The ratio of average noise intensity to average signal intensity is very high.

B) Similarity Measure

In image registration, the important criteria is the degree of similarity between the two images. The measure of similarity should be able to provide a large separation between the match point and all other test locations. Different values for the measure of similarity are shown in Fig. 3. We notice that the matching points (denoted by +) have lower distinct $A_D(i,j)$ values from those of the non-matched points.

C) The Sampling Rule

The rule to pick up the samples can affect the matching procedures. In Fig. 4, four different coarse spacings, randomly chosen, have been considered. Figures 5.a, 5.b, 5.c show the dependence of the measure of similarity $A_D(i,j)$ for different (i,j) 's on the sampling points. The figures demonstrate that block 1 converges faster. In block 1, each additional pixel is taken as far as possible from the original pixel.

D) Thresholding Sequence

From Fig. 3, we notice the clear distinction between the best match

AD-A131 923

ACQUISITION IMAGE AND DATA COMPRESSION(U) CITY COLL NEW
YORK D L SCHILLING ET AL. 30 APR 83 AFOSR-TR-83-0505
AFOSR-81-0169

2/2

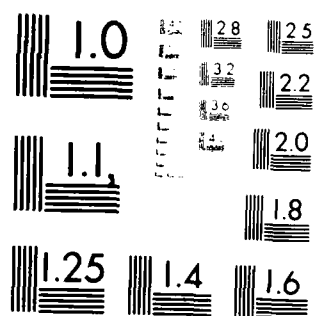
UNCLASSIFIED

F/G 17/2

NL



END
DATE
FILMED
9-83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

IV Performance comparison of the different criterion function

We have compared the performance of different measures of similarity. (These results have been shown in Table II. It turns out that the absolute difference measure is faster than other measures.)

These measures of similarity are as follows:

1. Average Absolute Difference Measure (ABS):

$$ABS(i,j) = \frac{1}{M} \sum_{p=1}^M \sum_{q=1}^M \left| S^{ij}(p,q) - T(p,q) \right|$$

2. Absolute Difference Measure (AB):

$$AB(i,j) = \frac{1}{4M} \sum_{p=1}^M \sum_{q=1}^M \left| S^{ij}(p,q) - T(p,q) \right|$$

3. Correlation Coefficient Measure (COE):

$$COE(i,j) = \frac{M^2 \sigma_2 - \sigma_4 \sigma_5}{\sqrt{M^2 \sigma_1 - \sigma_4^2} \sqrt{M^2 \sigma_3 - \sigma_5^2}}$$

where $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ and σ_5 as given in equation

4. Normalized Correlation Function (NCF):

$$NCF(i,j) = \frac{\sum_{p=1}^M \sum_{q=1}^M S^{ij}(p,q) T(p,q)}{\sum_{p=1}^M \sum_{q=1}^M [S^{ij}(p,q)]^2}$$

5. Direct Correlation Function (DCF):

$$DCF(i,j) = \frac{1}{M^2} \sum_{p=1}^M \sum_{q=1}^M S^{ij}(p,q) T(p,q)$$

IV Performance comparison of the different criterion function

We have compared the performance of different measures of similarity. (These results have been shown in Table II. It turns out that the absolute difference measure is faster than other measures.)

These measures of similarity are as follows:

1. Average Absolute Difference Measure (ABS):

$$ABS(i,j) = \frac{1}{M} \sum_{p=1}^M \sum_{q=1}^M \left| S^{ij}(p,q) - T(p,q) \right|$$

2. Absolute Difference Measure (AB):

$$AB(i,j) = \frac{1}{4M} \sum_{p=1}^M \sum_{q=1}^M \left| S^{ij}(p,q) - T(p,q) \right|$$

3. Correlation Coefficient Measure (COE):

$$COE(i,j) = \frac{M^2 \sigma_2 - \sigma_4 \sigma_5}{\sqrt{M^2 \sigma_1 - \sigma_4^2} \sqrt{M^2 \sigma_3 - \sigma_5^2}}$$

where $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ and σ_5 as given in equation

4. Normalized Correlation Function (NCF):

$$NCF(i,j) = \frac{\sum_{p=1}^M \sum_{q=1}^M S^{ij}(p,q) T(p,q)}{\sum_{p=1}^M \sum_{q=1}^M [S^{ij}(p,q)]^2}$$

5. Direct Correlation Function (DCF):

$$DCF(i,j) = \frac{1}{M^2} \sum_{p=1}^M \sum_{q=1}^M S^{ij}(p,q) T(p,q)$$

The results of the comparison between the five measures of similarity are shown in Table II. It turns out that the ABS measure gives the best match between the template and the scene. The COE is time consuming. The thresholding sequence $f(b)$ has been used as in reference (1), where $f(b)$ is an empirically derived function.

$$f(b) = (b^2 - 0.25b - 0.25) / b(b + 1)$$

The NCF measure is slightly slower than the COE measure. DCF function is the worst.

V Conclusion

In this paper a fast algorithm for image matching has been studied. The image similarity measure used is the sum of the absolute intensity difference between the template and the scene. The thresholding sequence has been found to be a constant and equals approximately 50% of the average intensity per pixel.

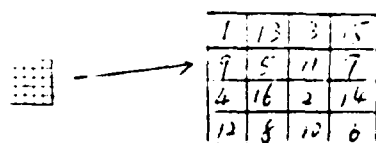
The performance of the algorithm has been tested by computer generated images (English letters) and Gaussian noise.

Comparison of the algorithm with other measures of similarity proves that the average absolute intensity difference is faster than others. We intend to investigate the algorithm performance with real images and at real time.

REFERENCES

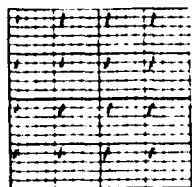
- (1) D.R. Sullivan and J.S. Martin, "A coarse search correlation tracker for image registration," IEE Trans. Aerosp. Electron. Syst., vol. AES-17, pp. 29-34, Jan. 1981.
- (2) D.I. Barnea and H.E. Silverman, "A class of algorithms for fast digital image registration," IEEE Trans. Comput. vol. CP-21, pp. 176-186, Feb. '72.
- (3) M. Svedlow, C. McGillem and P.E. Anuta, "Image registration: similarity measure and preprocessing method comparisons," IEEE Trans. Aerosp. Electron. Syst., Vol. AES-14, pp. 141-149, Jan. 1978.
- (4) E.L. Hall, D.L. Davies and ME. Casey, "The selection of critical subsets for signal image, and scene matching," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. PAMI-2, pp. 313-322, July 1980.
- (5) R.Y. Wong and E.L. Hall, "Performance comparison of scene matching techniques," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. PAMI-1, pp. 325-330, July 1979.
- (6) R.Y. Wong, "Sequential scene matching using edge features," IEEE Trans. Aerosp. Electron Syst., Vol. AES-14, January 1978.
- (7) H.K. Ramapriyan, "A multilevel approach to sequential detection of pictorial features," IEEE Trans. Comput., Vol. C-25, pp. 66-78, Jan., '76.
- (8) R.Y. Wong, E.L. Hall, and J.D. Rouge, "Hierarchical search for image matching," in 1976 Proc. IEEE Conf. Decision and Control, Dec. 1976.

Block 1



1	13	3	5
9	5	11	7
4	16	2	14
12	8	10	6

subset b=1



subset b=2



subset b=3



....

Fig. 1 Sampling manner of the data set
 (each pass requires that one additional pixel be chosen from each block)

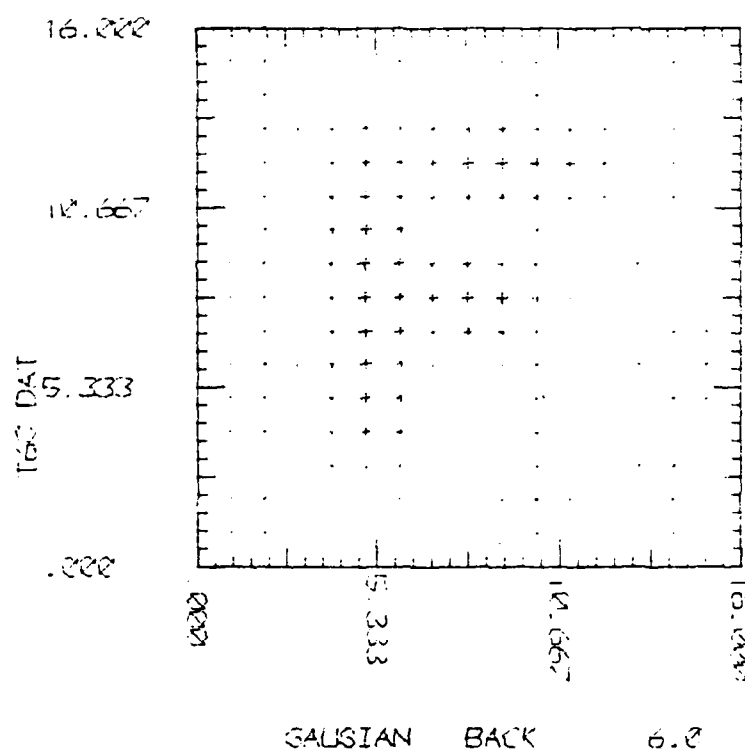


Fig.1d. Template image

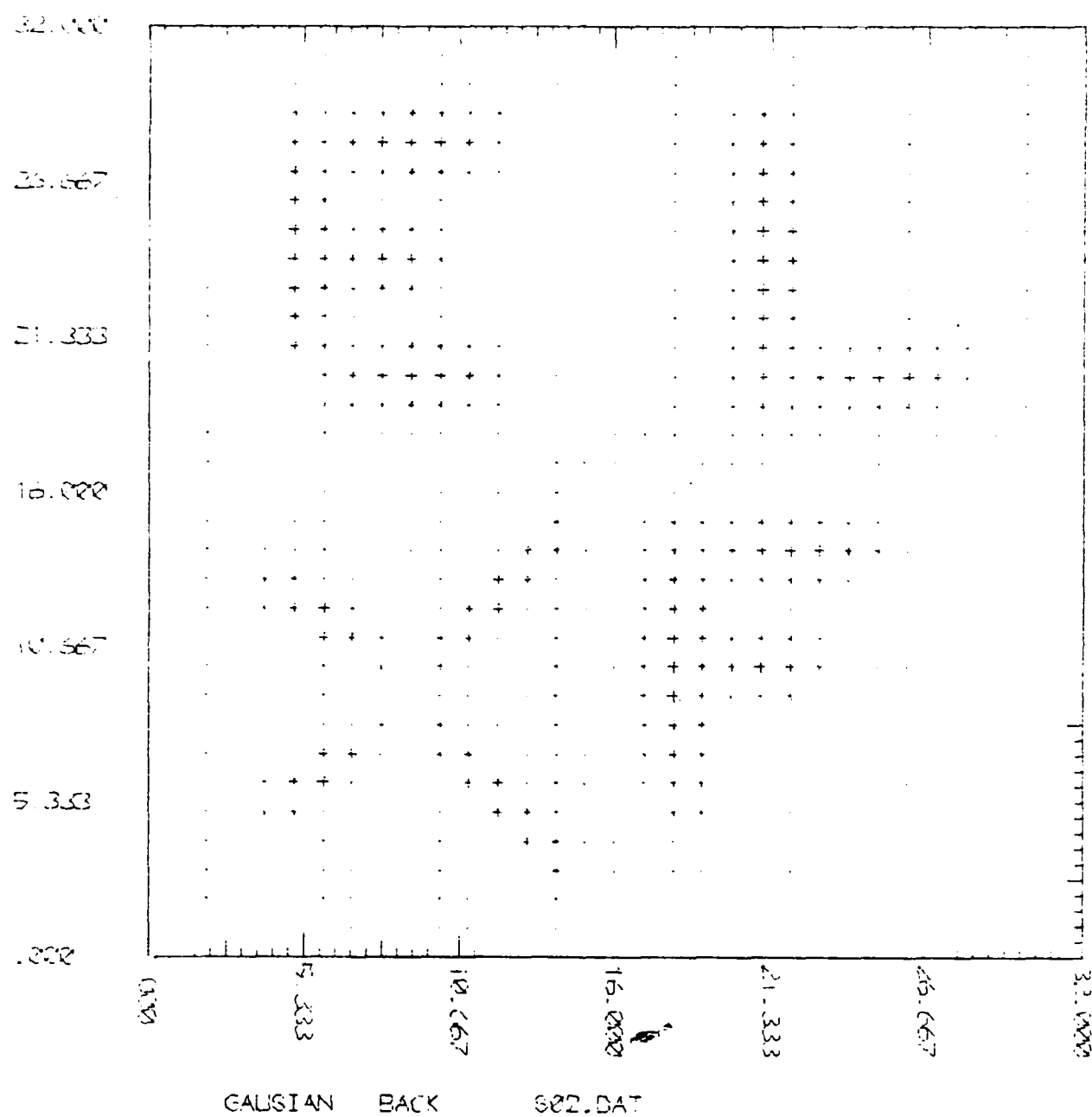


Fig.2b the typical scene image

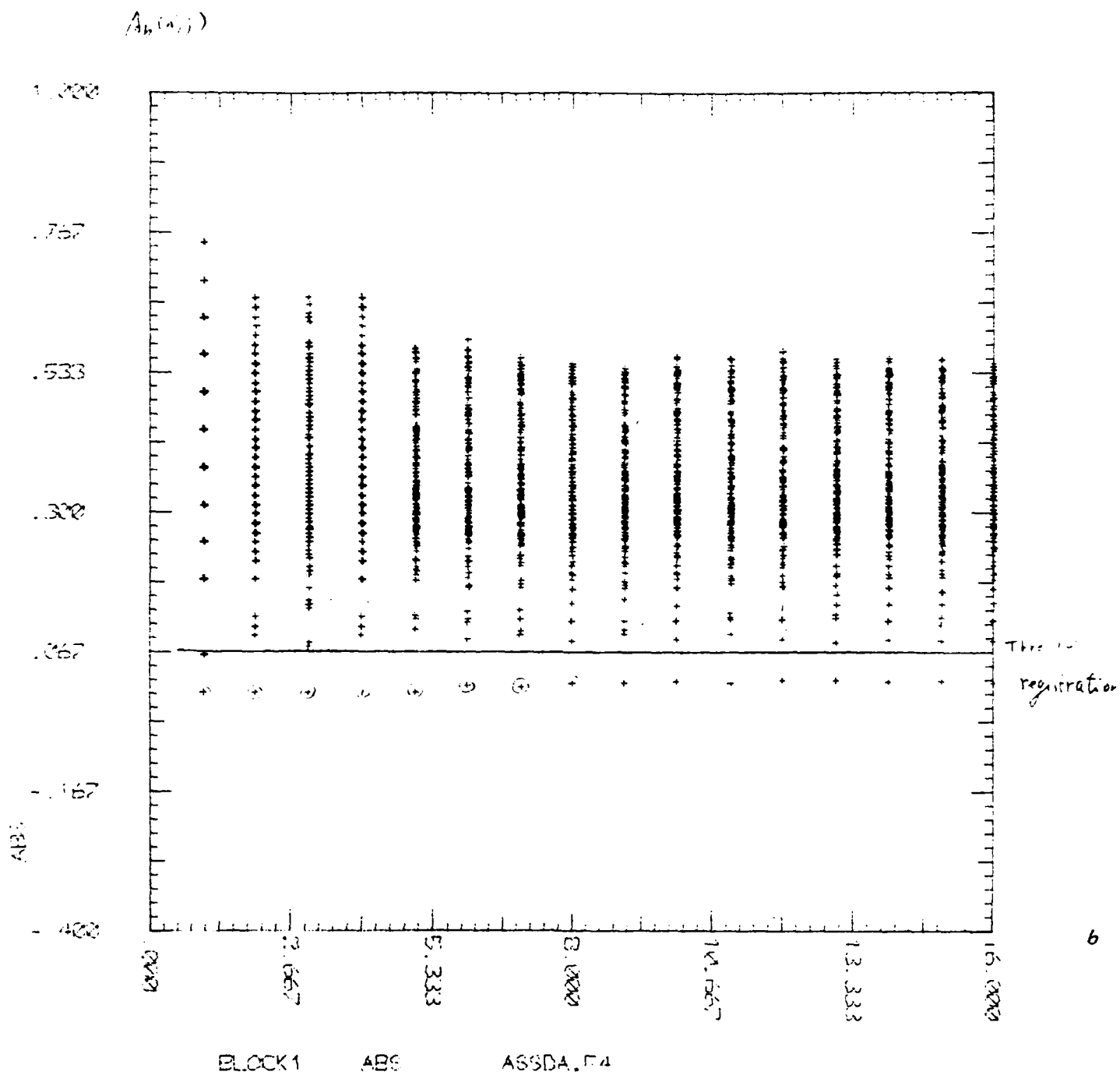


Fig. 3. The statistical behavior of the $A_b(i,j)$.

Block 1

1	13	3	15
9	5	11	7
4	16	2	14
12	8	10	6

Block 2

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

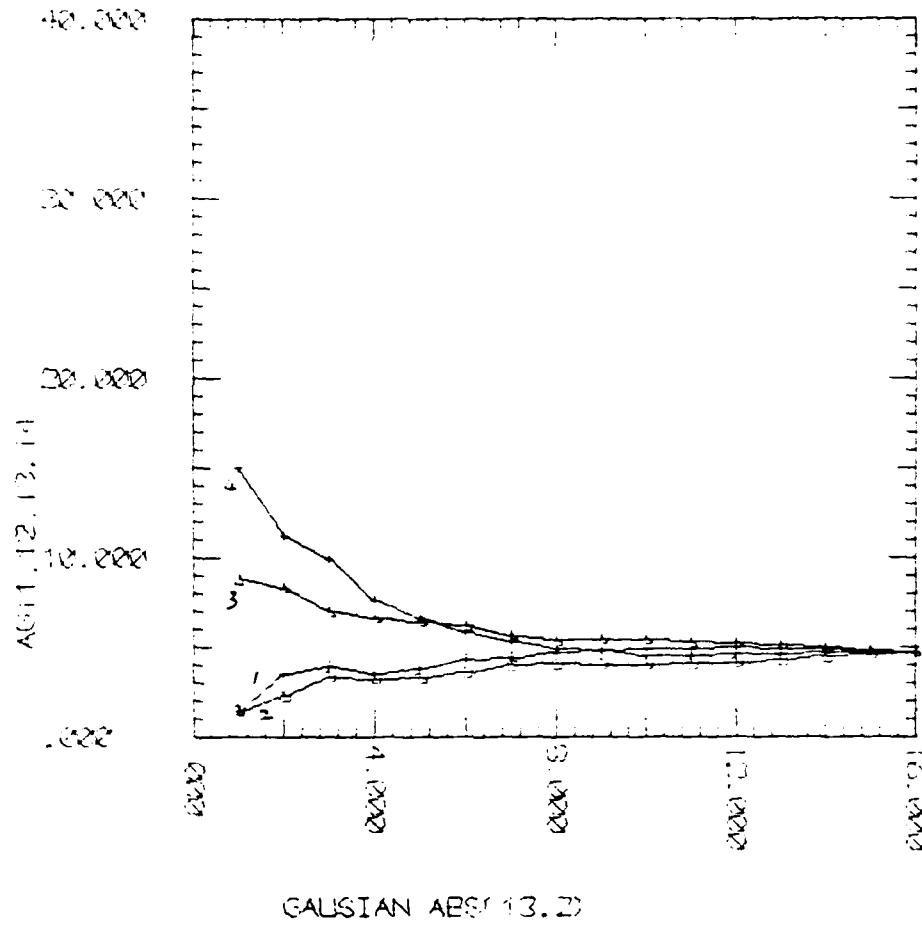
Block 3

13	14	11	16
5	12	17	12
1	6	1	8
1	2	3	4

Block 4

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----

Fig.4. Sampling manners

$\lambda(3,2)$ 

6

Fig. 5a.

BLOCK 1,2,3,4 ABS(13,2)

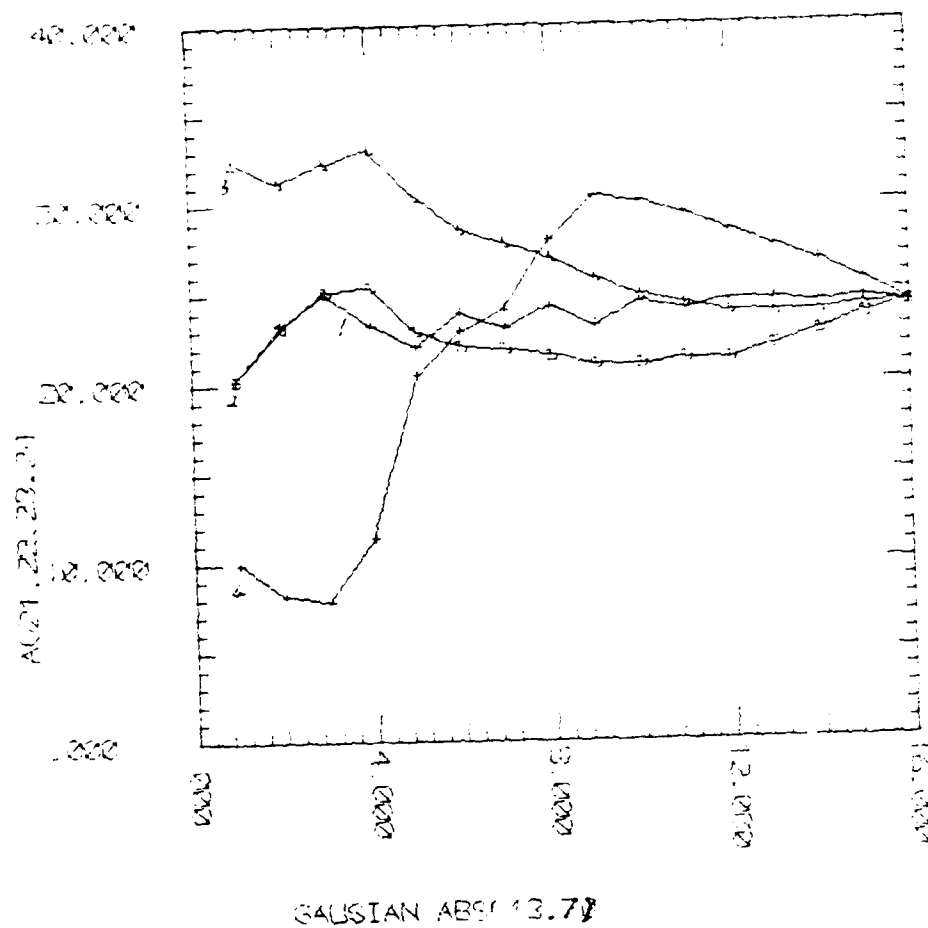
$\Delta_n(13, 7)$


Fig. 5b.

BLOCK 1,2,3,4, ABS(13,7)

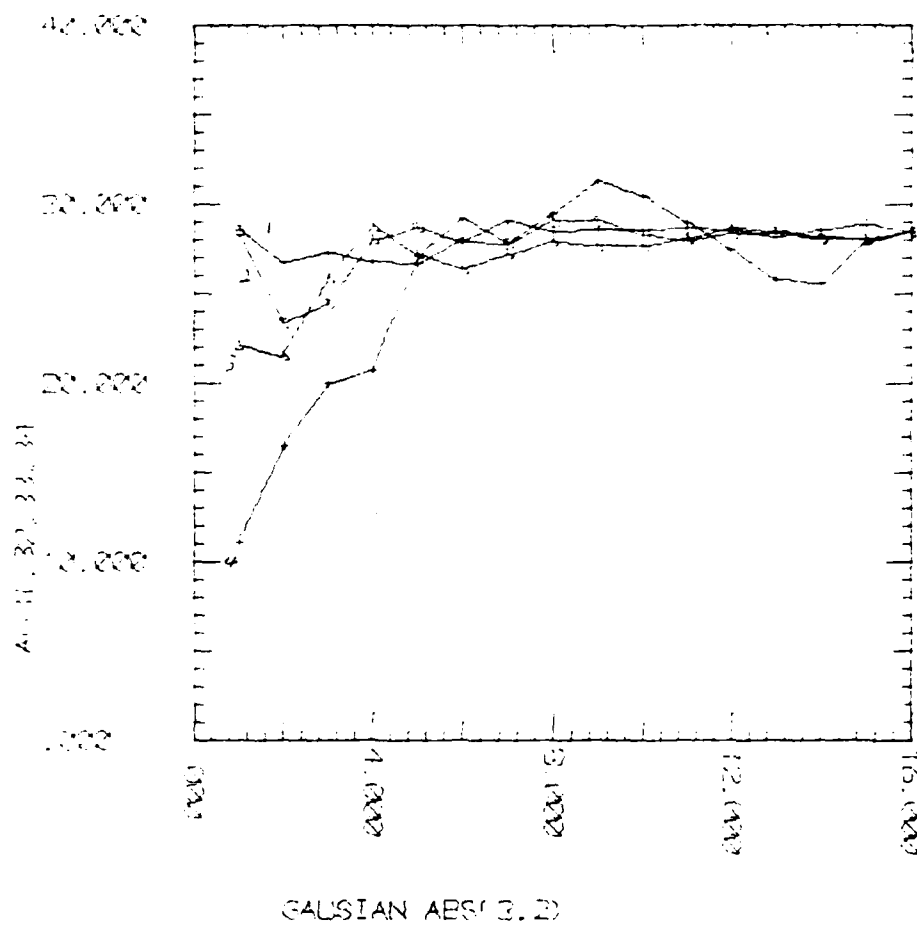
$A_0(3,1)$


Fig. 3c.

BLOCK 1,2,3,4, ABS(3,2)

**DAT
FILM**